

Optimized Dynamical Decoupling via Genetic Algorithms

Gregory Quiroz

*Department of Physics and Center for Quantum Information Science & Technology,
University of Southern California, Los Angeles, California 90089, USA*

Daniel A. Lidar

*Departments of Electrical Engineering, Chemistry, and Physics,
and Center for Quantum Information Science & Technology,
University of Southern California, Los Angeles, California 90089, USA*

We utilize genetic algorithms to find optimal dynamical decoupling (DD) sequences for a single-qubit system subjected to a general decoherence model under a variety of control pulse conditions. We focus on the case of sequences with equal pulse-intervals and perform the optimization with respect to pulse type and order. In this manner we obtain robust DD sequences, first in the limit of ideal pulses, then when including pulse imperfections such as finite pulse duration and qubit rotation (flip-angle) errors. Although our optimization is numerical, we identify a deterministic structure underlies the top-performing sequences. We use this structure to devise DD sequences which outperform previously designed concatenated DD (CDD) and quadratic DD (QDD) sequences in the presence of pulse errors. We explain our findings using time-dependent perturbation theory and provide a detailed scaling analysis of the optimal sequences.

I. INTRODUCTION

Quantum information processing (QIP) relies on the ability to implement high-fidelity quantum gate operations and successfully preserve quantum state coherence [1]. One of the most challenging obstacles for reliable QIP is overcoming the inevitable interaction between a quantum system and its environment or bath. Unwanted interactions result in decoherence processes that cause quantum states to deviate from a desired evolution, consequently leading to computational errors and loss of coherence. In order for QIP to be realizable in the setting of open quantum systems, it is necessary to address the detrimental effects of decoherence.

Dynamical decoupling (DD) is one such method, which seeks to attenuate the effects of decoherence by applying strong and expeditious control pulses solely to the system [2–7]. Provided the pulses are applied over a time duration sufficiently shorter than the correlation time associated with the environment dynamics, DD effectively averages out undesirable interactions and preserves quantum states with a low probability of error, or fidelity loss. One advantage of DD over quantum error correction (QEC) is that it is an open-loop technique, i.e., does not require feedback or measurement. Furthermore, DD has been widely experimentally studied in a number of systems, including ion traps [8–10], nuclear magnetic resonance (NMR) [11–13], solid state quantum dots [14], and nitrogen vacancy (NV) centers in diamond [15–17].

The earliest known DD sequence constructions built upon the Hahn spin echo effect [18], by applying a sequence of control pulses implementing π -rotations, each separated by a fixed time duration. Two particularly notable sequences include CPMG [19], which utilizes cycles of two identical pulses to preserve spin magnetization along a single direction (useful for known quantum state preservation), and XY₄ [20], a four pulse multi-axis DD sequence that can increase coherence times isotropically (useful for unknown quantum state preservation). Both sequences were eventually extended to

an “XY-family” which incorporates longer pulse sequences to improve coherence even further and provide robustness against errors generated by experimental imperfections in the control pulses [21].

In the context of open quantum systems of interest to us here, the key feature responsible for an increase in coherence time for CPMG and XY₄ is the suppression of the first order term in time-dependent perturbation theory for a system weakly coupled to a bath and subjected to dephasing or general decoherence, respectively. Concatenated DD (CDD) is a deterministic sequence design that exploits this property by recursively embedding any base DD sequence (e.g., XY₄) into itself to successively suppress an additional order of the perturbation expansion at each level of concatenation [22]. CDD has been extensively studied analytically [23–25], numerically [26–28], and experimentally [14, 29–33], for a variety of systems, and its predicted ability to achieve high order suppression has been largely confirmed.

A CDD sequence of order q using a base sequence of K pulses uses K^q equally spaced pulses to suppress the first q orders of the perturbation expansion in the ideal pulses limit for a single qubit system subjected to general decoherence (e.g., $K = 4$ for XY₄) [22, 23]. What distinguishes an XY₄-based CDD sequence of order q from, e.g., a periodically repeated XY₄ sequence (PDD) with the same total number of pulses, is just the pulse order. Yet, (unsymmetrized) PDD achieves at most first order suppression, in contrast to q th order suppression for CDD. This begs the question of whether there exist other sequences that achieve even better performance than CDD, arrived at merely by optimizing the pulse order and type. This is the main question we address in this work, using numerical optimization based on genetic algorithms, supplemented with simulated annealing.

Notwithstanding recent progress in pulse sequence optimization using unequal intervals, in particular the UDD sequence and its generalizations [34–44], we focus our optimization on the case of equal pulse-intervals, as this greatly simplifies the optimization, and is furthermore often the most

convenient experimental situation. Our first main result is to show that numerically optimal sequences comprising ideal, zero-width π -pulses can be located which exhibit the same error suppression characteristics as XY_4 -based CDD while using 2^{-r} times fewer pulses, where r is the concatenation level. Furthermore, we identify in these sequences a deterministic structure which we use to construct additional, robust high-order sequences.

As the techniques of DD sequence construction have become increasingly sophisticated, the issue of robustness to control pulse imperfections has remained one of the prominent restrictions of sequence performance in experimental settings. Systematic errors such as rotation-angle or rotation-axis errors, and finite pulse duration errors brought about by bandwidth constraints, can generate additional decoherence that quickly destroys the decoupling efficiency of all known DD schemes. Robustness to such errors has been addressed by the XY-family of sequences and CDD, and in a systematic manner—for pulse-width errors—by Eulerian DD (EDD) [45] and its generalization to logic gates, known as dynamically corrected gates (DCG) [46, 47]. A concatenated version of DCG (CDCG) has been shown to be capable in principle of achieving arbitrarily accurate gates using finite width pulses [48]. Protocols based on pulse-interval optimization have been shown experimentally to be highly sensitive to pulse imperfections, thus forfeiting their ideal-pulse decoupling efficiencies [49]. Certain numerical optimization techniques such as locally optimized DD (LODD) [50], bandwidth adapted DD (BADD) [51], optimized noise filtration DD (OFDD) [52], and Walsh function DD (WDD) [53], exhibit a degree of robustness to finite pulse duration, however, the relationship between sequence performance and rotation errors is unclear. A more recent approach for combating pulse imperfections, known as Knill DD (KDD), utilizes a sequence of variable angle pulses separated by fixed pulse-intervals to generate an effective sequence of two π -pulses [54]. In contrast to the XY-family and XY_4 -based CDD, KDD exhibits robustness to finite-width and flip-angle errors. One important question which we seek to address here is whether it is possible to obtain a similar robustness to both forms of pulse imperfections utilizing *only* π -pulses.

To account for such errors, our numerical optimization is extended to include pulse imperfections, in particular finite pulse width, flip-angle errors, or both. We show that robust DD sequences exist which perform considerably better than the original CDD sequence and sequences based on unequal pulse-intervals, such as UDD and its variants [34–44]. However, sequence performance eventually saturates with growing sequence length. Interestingly, we find that the deterministic structure identified in the ideal pulse limit provides robustness against both forms of pulse errors.

The structure of this paper is as follows. In Section II, we supply background information and formal mathematical specifications for DD effectiveness in the context of general open quantum system dynamics. The particular error model utilized for this study is then discussed along with the details regarding the control Hamiltonian, which is responsible for the DD control fields. In Section III, we discuss how to apply

Genetic Algorithms (GAs) to DD optimization. The results of our optimal sequence search are then given in Section IV, where we identify sequences that obtain a higher degree of DD efficiency than standard CDD sequences under the condition of ideal δ -function pulses, and exhibit robustness to certain forms of pulse imperfections; namely, over and under-rotation errors. In Section V, the search results are supplemented with a comparison between the GA optimal sequences, CDD, and QDD for each pulse specification defined in Sec. II. Optimization proves to be most notably beneficial in the case of pulse imperfections where the GA sequences convey their superiority over known deterministic sequences. In Sec. VI, using the results of Sec. IV and V, we discuss the existence of a concatenation-based deterministic scheme built from optimal sequences that offers robustness to both finite-width and rotation errors. We present our conclusions in Section VII, while additional information regarding DD performance scaling is given in Appendices A and B.

II. BACKGROUND, PROBLEM SETUP, AND TOOLS

In this section, we briefly review the pertinent basic background and mathematical framework for DD. The general error model is then introduced and its numerical implementation specified. The DD control Hamiltonian is discussed and specified for each of the four pulse conditions employed in this work, differing by the degree and nature of the pulse imperfections. Finally, we introduce a distance measure with which DD performance is quantified throughout the paper.

A. Dynamical Decoupling

Consider an open quantum system described by the Hamiltonian

$$H(t) = H_0 + H_C(t). \quad (1)$$

The time-independent term H_0 governs the internal dynamics of the system and environment, while $H_C(t)$ is responsible for the time-dependent DD control fields. The Hamiltonian H_0 is resolved further into

$$H_0 \equiv H_S \otimes I_B + I_S \otimes H_B + H_{SB}, \quad (2)$$

where H_S is the pure system Hamiltonian, H_B is the pure environment Hamiltonian, H_{SB} represents the system-environment interaction, and $I_{S(B)}$ is the identity operator on the system (bath).

For brevity, we denote

$$H_{\text{err}} \equiv H_S \otimes I_B + H_{SB} \quad (3)$$

as the *error* Hamiltonian, where the pure system and system-environment interaction Hamiltonians constitute the sources of undesired system evolution and decoherence, respectively. Removal of undesired system evolution is particularly relevant when DD is utilized for high fidelity quantum memory

storage [2, 23, 34] or in a “decouple-then-compute” approach to quantum gate construction [25, 55, 56]. Storing quantum memory requires initial state preservation, hence the desired system evolution is trivial action on the system. Any form of system dynamics present after the DD evolution would alter the initial state, resulting in storage errors. In a similar manner, gate errors can be acquired during the application of a nontrivial quantum gate if undesired system dynamics remain upon the completion of the gate operation. Undesired system evolution must also be removed in alternative DD-protected gate construction strategies, such as “decouple-while-compute” [55–57], in particular to prevent leakage errors when a decoherence-free subspace (DFS) or stabilizer code is used to enable computation while DD pulses are applied [58, 59].

We assume that the control Hamiltonian $H_C(t)$ implements the DD pulses solely on the system. $H_C(t)$ consequently acts trivially on the pure environment Hamiltonian, $[H_C(t), H_B] = 0 \forall t$, and nontrivially on H_{err} . The manner in which $H_C(t)$ operates on H_{err} ultimately determines the effectiveness of DD in suppressing the contributions of the error Hamiltonian to the system evolution. Demanding that each pulse operator anticommute with at least one term comprising H_{err} , the system is driven in such a way that H_{err} can be effectively averaged out for sufficiently short time durations. How short can be elucidated in the interaction (“toggling”) picture with respect to $H_C(t)$ [3, 55, 60, 61], where

$$\tilde{H}_0(t) = U_C^\dagger(t) H_0 U_C(t) \quad (4a)$$

$$= \tilde{H}_{\text{err}}(t) + I_S \otimes H_B \quad (4b)$$

with the control unitary

$$U_C(t) = \mathcal{T} \exp \left(-i \int_0^t H_C(t) dt \right), \quad (5)$$

where \mathcal{T} denotes the time ordering operator. The unitary time evolution operator $\tilde{U}_0(t)$ satisfies the Schrödinger equation

$$i \frac{\partial}{\partial t} \tilde{U}_0(t) = \tilde{H}_0(t) \tilde{U}_0(t), \quad \tilde{U}_0(0) = I \quad (6)$$

and $\tilde{U}_0(t) = U_C^\dagger(t) U(t)$, where $U(t)$ is the time evolution operator generated by Eq. (1).

Employing time-dependent perturbation theory (TDPT) via the Magnus expansion [62, 63] to solve Eq. (6) we can write

$$\tilde{U}_0(\tau_c) = \exp \left(\sum_{n=1}^{\infty} \Omega^{(n)}(\tau_c) \right), \quad (7)$$

with the anti-Hermitian operator $\Omega^{(n)}(\tau_c)$ representing the n th term in the Magnus operator expansion after a total DD cycle time τ_c . The leading terms of the expansion are

$$\Omega^{(1)}(\tau_c) = -i \int_0^{\tau_c} \tilde{H}_0(t_1) dt_1, \quad (8)$$

$$\Omega^{(2)}(\tau_c) = -\frac{1}{2} \int_0^{\tau_c} dt_1 \int_0^{t_1} dt_2 [\tilde{H}_0(t_1), \tilde{H}_0(t_2)], \quad (9)$$

while the higher order Magnus terms are constructed recursively as a sum of $(n-1)$ -fold commutators. A sufficient condition for convergence of the Magnus expansion is [64]

$$\int_0^{\tau_c} \|\tilde{H}_0(t)\| dt < \pi. \quad (10)$$

In agreement with average Hamiltonian theory (AHT) [60, 65], the time-dependent evolution generated by $\tilde{H}_0(t)$ is formally identical to a time-independent evolution generated by the effective Hamiltonian

$$\begin{aligned} \bar{H}_0 &= \frac{i}{\tau_c} \sum_{n=1}^{\infty} \Omega^{(n)}(\tau_c) \\ &= \bar{H}_B(\tau_c) + \bar{H}_{\text{err}}(\tau_c). \end{aligned} \quad (11)$$

AHT applies here since we are only interested in the joint system-environment dynamics at the end of each stroboscopic DD period of τ_c . We can further partition \bar{H}_0 into an effective pure environment term and a sum of effective error Hamiltonians

$$\bar{H}_{\text{err}}(\tau_c) \equiv \frac{i}{\tau_c} \sum_{n=1}^{\infty} \Omega_{\text{err}}^{(n)}(\tau_c), \quad (12)$$

where $\Omega_{\text{err}}^{(n)}(\tau_c)$ is the n th order Magnus operator containing non-trivial system operators, while $\bar{H}_B(\tau_c)$ contains only terms with trivial action on the system.

In light of Eq. (12) we find that DD facilitates an effective suppression of H_{err} by suppressing $\bar{H}_{\text{err}}(\tau_c)$ up to some order in the Magnus expansion. When the first N terms of the expansion of $\bar{H}_{\text{err}}(\tau_c)$ vanish we speak of “ N th order decoupling.” Assuming N th order decoupling has been achieved, the toggling frame evolution is given by

$$\tilde{U}_0(\tau_c) = e^{-i\tau_c[H_B(\tau_c) + H_{\text{err}}(\tau_c)]} \quad (13a)$$

$$= e^{-i\tau_c \bar{H}_B(\tau_c) + \mathcal{O}[(\|\bar{H}'_{\text{err}}\| \tau_c)^{N+1}]}, \quad (13b)$$

where the evolution is predominately dictated by the effective pure environment Hamiltonian $\bar{H}_B(\tau_c)$ when $\|\bar{H}'_{\text{err}}\| \tau_c \ll \pi$ and $N \gg 1$. Here, $\bar{H}'_{\text{err}} = \frac{i}{\tau_c} \sum_{n=N+1}^{\infty} \Omega_{\text{err}}^{(n)}(\tau_c)$ denotes the remaining effective error Hamiltonian and $\|A\|$ is the sup-operator norm of A (largest singular value):

$$\|A\| = \sup_{|\psi\rangle} \frac{\|A|\psi\rangle\|}{\| |\psi\rangle \|}. \quad (14)$$

Thus the effectiveness of DD is dependent upon intrinsic properties, in particular the strength of the interaction and pure environment Hamiltonians. In situations where the internal dynamics are sufficiently fast, a short DD cycle time is desirable to maintain Eq. (10). Furthermore, it is also desirable to achieve high order error suppression, $N \gg 1$, so that the effects of \bar{H}'_{err} are less consequential. We make use of both conditions to analyze DD in the presence of various strengths of internal dynamics, while determining the minimum number of control pulses required to obtain a given order of error suppression.

As we shall see when we present the analysis of optimal sequences, starting in Sec. IV, the effective error Hamiltonian (12) has significant explanatory power.

B. Error Model

The system of interest is a single-qubit system generically coupled to its environment. The internal dynamics are governed by

$$H_0 = \sum_{\mu \in \{I, x, y, z\}} \sigma^\mu \otimes B_\mu, \quad (15)$$

where σ^μ and B_μ are the spin-1/2 Pauli matrices and general bounded environment operators, respectively. Selecting a four-qubit spin bath to model the environment for the numerical search, the environment operators are given by

$$B_\mu = \sum_{i \neq j} \sum_{\alpha, \beta} c_{\alpha\beta}^\mu \left(\sigma_i^\alpha \otimes \sigma_j^\beta \right), \quad (16)$$

where i, j index the bath qubits, $\alpha, \beta, \mu \in \{I, x, y, z\}$, and $c_{\alpha\beta}^\mu \in [0, 1]$ are random coefficients chosen from a uniform probability distribution. The construction of B_μ permits at most two-body interactions between the environment qubits and three-body interactions between the system and environment. Note that Eq. (16) contains terms proportional to the identity operator $\sigma_i^I \otimes \sigma_j^I$, which account for the pure system Hamiltonian described in Eqs. (2) and (3).

C. Control Hamiltonian

The general form of a single qubit control Hamiltonian is

$$H_C(t) = \frac{1}{2} \sum_{\mu \in \{x, y, z\}} V_\mu(t) \sigma^\mu, \quad (17)$$

where $V_\mu(t)$ is the control field associated with the σ^μ degree of freedom. All of the essential information regarding the DD sequence is contained within $V_\mu(t)$, i.e., pulse timings and amplitude profiles. In general, varying either quantity can result in drastically different optimal sequence constructions. Considering equal pulse-interval delay times of τ_d throughout the DD evolution, we examine how optimal sequence construction varies with amplitude profile. We consider the most customary pulse profiles: zero-width and rectangular, finite-width, and pursue an analysis of each profile with the addition of qubit rotation errors to model the existence of systematic errors brought about by faulty control fields.

1. Ideal pulses

The first type of pulse considered is an idealized, zero-width control field

$$V_\mu(t) = \sum_j \phi_0 \delta(t - t_j^\mu), \quad (18)$$

where the Dirac delta function $\delta(t)$ constitutes the pulse profile. The pulses are applied at times t_j^μ and the angle of rotation is given by ϕ_0 . The control fields are restricted so that

they act uni-axially for all time t . In terms of the single-qubit Bloch sphere, the condition can be visualized as allowing pulses solely along one of the three axes. We impose this constraint on all subsequent definitions of $V_\mu(t)$ as well.

2. Finite-width pulses

Since zero-width pulses are experimentally impossible, we relax the ideal pulse assumption and consider pulses of finite duration as well. We model the finite duration by

$$V_\mu(t) = \sum_j A [\Theta(t - t_j^\mu) - \Theta(t - t_j^\mu - \tau_p)], \quad (19)$$

representing a piecewise continuous control field with a rectangular profile [2]. The pulse amplitude is denoted by A and the pulse duration is τ_p , so that $A\tau_p = \phi_0$. The Heaviside Theta function, $\Theta(t)$, dictates the pulse profile where it is assumed that the time to turn the pulse “on” and “off” is negligible, therefore, the pulse is well approximated by a square wave.

3. Flip-angle errors

An additional form of systematic error we consider is that of an over- or under-rotation in the angle ϕ_0 , commonly referred to as a flip-angle error [60]. This particular type of error is relevant, e.g., in nuclear magnetic resonance (NMR), where inhomogeneity of the control field across the sample results in qubit rotation errors [12]. Flip-angle errors are also prevalent in other systems such as donor electron spins in Si systems [66, 67].

In the case of zero-width pulses, the control field takes the form

$$V_\mu(t) = \sum_j \phi_0 (1 \pm \epsilon) \delta(t - t_j^\mu), \quad (20)$$

where ϵ denotes the error in the rotation angle and the $+$ ($-$) refers to an over-(under-)rotation. By modeling the control field in this manner, it is assumed that the pulses are applied along their respective axes with zero or negligible error. The inclusion of rotation-axis errors has been previously studied for some common deterministic DD schemes [68], but is not included in our present study.

4. Finite-width flip-angle errors

As a worst case scenario we also consider the combined effect of flip-angle errors for finite-width pulses. Assuming that the error in the pulse duration is negligible, a flip-angle error can be thought of as an error in the pulse amplitude. We model the combined error control field by

$$V_\mu(t) = \sum_j A (1 \pm \epsilon) [\Theta(t - t_j^\mu) - \Theta(t - t_j^\mu - \tau_p)] \quad (21)$$

and note that this particular form is one of the most prevalent pulse profiles encountered in experimental settings [13, 49, 54, 66, 68].

D. Distance Measure and Scaling

Rather than the standard Uhlmann fidelity or trace-norm distance [1] we use a state-independent distance measure, which significantly reduces the computational overhead. Namely, we quantify DD performance using

$$D(U, G) = \frac{1}{\sqrt{2d_S d_B}} \min_{\Phi} \|U - G \otimes \Phi\|_F, \quad (22)$$

where U represents the full evolution operator of the sequence [i.e. U satisfies the Schrödinger equation (6) with the Hamiltonian (1)], G is the desired evolution of the system, the norm is the Frobenius norm $\|X\|_F = \sqrt{\text{Tr}(X^\dagger X)}$, and d_S and d_B are the dimensions of the system and the environment Hilbert spaces \mathcal{H}_S and \mathcal{H}_B , respectively [69]. The minimization problem can be solved analytically to obtain the closed form expression [70]:

$$D(U, G) = \sqrt{1 - \frac{1}{d_S d_B} \|\Gamma\|_{\text{Tr}}}, \quad (23)$$

where $\|\Gamma\|_{\text{Tr}} = \text{Tr}(\sqrt{\Gamma^\dagger \Gamma})$ represents the trace-norm and

$$\Gamma = \text{Tr}_S[U(G^\dagger \otimes I_B)], \quad (24)$$

where Tr_S denotes a partial trace over the system degrees of freedom. In the subsequent analysis, $G \equiv I_S$ for the desired DD evolution and we denote $D \equiv D(U, I)$.

We characterize optimal sequence performance with respect to two parameters associated with the internal dynamics: the “strength” of the error Hamiltonian and pure environment dynamics given by

$$J = \|H_{\text{err}}\|, \quad \beta = \|H_B\|, \quad (25)$$

respectively, and three parameters whose relevance depends on the control Hamiltonian specifications: the pulse-interval τ_d , the pulse duration τ_p , and the rotation angle error ϵ . Each of these parameters can be utilized to extract the scaling of the dominant term in the effective Hamiltonian by analyzing D as a function of the parameter of interest.

In the case of ideal pulses, the distance can be shown to be upper-bounded as

$$D \lesssim \mathcal{O}[(J + \beta)^{N+1} \tau_c^{N+1}], \quad (26)$$

where N is the order of error suppression. See Appendix A for a proof. Analyzing D as a function of τ_d , J , and β , the order of error suppression and essentially the structure of the dominant effective error Hamiltonian operator can be determined for the relevant situations where the dynamics are dominated by system-environment interactions ($J \gg \beta$) or the bath dynamics ($J \ll \beta$). Similar studies can be performed for finite-width or flip-angle errors as well. We ultimately utilize this method in conjunction with AHT to characterize each optimal sequence and determine robustness to various pulse errors.

III. ALGORITHM

Genetic Algorithms represent an approach to optimization problems based on the properties of natural evolution. Given an initial population and a definition of fitness, the algorithm simulates the processes of selection, reproduction, and mutation in an attempt to locate the member in the population with the highest probability of survival. In regards to DD, the population can be thought of as a subset of all possible sequence configurations, where a configuration is specified by the order and types of pulses, for a given sequence length K . The member with the highest probability of survival is the sequence which maximally suppresses system-bath interactions with respect to a particular distance measure. In the following sub-sections we outline the representation of the population and discuss how selection, reproduction, and mutations are implemented in the setting of DD optimization.

A. Sequence construction

The search is performed in the “lab” reference frame defined by the Hamiltonian $H(t)$ of Eq. (1). Due to the piecewise continuous form of $H_C(t)$, the time evolution operator $U(\tau_c)$ can be decomposed into a series of unitary operators. Demanding fixed pulse-intervals, a sequence of K pulses applied at $t_j = j\tau_d$, $j = 1, 2, \dots, K$, yields the time evolution operator

$$U(\tau_c) = P_K f_{\tau_d} P_{K-1} f_{\tau_d} \cdots P_2 f_{\tau_d} P_1 f_{\tau_d}, \quad (27)$$

where P_j is the unitary evolution operator achieved by the j th pulse and $f_{\tau_d} = e^{-iH_0\tau_d}$ designates the “free evolution” propagator between successive pulses. The pulse operators are defined such that $P_j \in \mathcal{G}$, where \mathcal{G} denotes a discrete set of allowable control pulses that depends upon the choice of $V_\mu(t)$. The total sequence time τ_c is also dictated by the choice of $V_\mu(t)$ since the finite duration of the pulse contributes when applicable, e.g., for a sequence of m_d pulse delays and m_p nontrivial pulses $\tau_c = m_d\tau_d + m_p\tau_p$. While Eq. (27) is not the most general DD evolution operator for fixed pulse-intervals, since it does not permit consecutive pulses without free evolution periods, it still captures a majority of the known sequences and additional highly robust sequence constructions.

B. Chromosome structure

The canonical approach to GAs is to define a member of the population by a set of genes, loosely referred to as a chromosome. Each gene can be thought of as a parameter in the optimization problem which contributes in some way to the fitness, and therefore the probability of selection, of the member. Defining a member in the population as a DD sequence, Eq. (27) is translated directly into its corresponding chromosome

$$C_j^{(\alpha)} = \{P_1, P_2, \dots, P_{K-1}, P_K\}, \quad (28)$$

representing the j th member in the α th generation. The genes are given by the pulses in the sequence, therefore the number of genes increases with increasing sequence length. In general, a sequence and its corresponding chromosome do not have to be structurally equivalent. Later we will elaborate on why the naive translation of Eq. (28) is not favorable for DD optimization and discuss how it can be refined; however, for now Eq. (28) is adequate to describe each aspect of the algorithm outlined in the subsequent sub-sections.

The population is given by the set of chromosomes $\{C_j^{(\alpha)}\}_{j=1}^Q$, where each $C_j^{(\alpha)}$ corresponds to a sequence $U_j^{(\alpha)}(\tau_c)$ and Q is the population size. The total number of possible sequence configurations, $\mathcal{N}(K)$, is determined by both the length of the sequence and the number of pulse types in \mathcal{G} . The size of the sequence space grows exponentially with the length of the sequence, $\mathcal{N}(K) = |\mathcal{G}|^K$, where $|\mathcal{G}|$ is the number elements in \mathcal{G} .

The search space can be greatly reduced by imposing the cyclic DD condition $U_C(\tau_c) = I_S$, which is applicable for our focus on quantum memory preservation. The condition can be recast in the context of the search problem as

$$\prod_{j=1}^K P_j^{\text{ideal}} \sim I_S \quad (29)$$

on all $C_j^{(\alpha)}$, where only the ideal, zero-width version of the pulse is used when finite-width or flip-angle error pulse profiles define $V_\mu(t)$. Applying Eq. (29), the search space is reduced to $\mathcal{N}_R(K) = |\mathcal{G}|^{K-1}$, where only $|\mathcal{G}|^{-1}$ of the original search space accounts for viable DD sequences.

The initial population is chosen at random from the reduced search space, such that $Q \ll \mathcal{N}_R(K)$. In general, the size of Q is somewhat arbitrary and expected to vary depending on the number of degrees of freedom specified by the problem. In the context of DD optimization, the size of the initial population will ultimately end up fixed for all sequence lengths due to the structure of the initial chromosomes; see Section III F 1 for additional details.

C. Selection

Associated with each chromosome $C_j^{(\alpha)}$ is a selection probability $p_j^{(\alpha)}$. This quantity defines the probability of being selected for reproduction in generation α and is given by

$$p_j^{(\alpha)} = \frac{q_j^{(\alpha)}}{\sum_i q_i^{(\alpha)}}, \quad (30)$$

where $q_j^{(\alpha)} = -\log_{10} D_j^{(\alpha)}$ represents the performance, or fitness, of the j th sequence. Here, we impose the cyclic DD condition as well, $G = I_S$, and denote $D_j^{(\alpha)} \equiv D_j^{(\alpha)}(U(\tau_c), I_S)$. The logarithm is included in the definition of the fitness due to complications with the selection probability that are attributed to the extreme sensitivity of Eq. (22), and any distance measure for that matter, to sequence variations. The exchange of

a single pulse in a sequence with any other member of the decoupling set can result in a change in performance up to many orders of magnitude. Since the reduced search space does not dismiss all poor-performing sequences, the fitness can range greatly in any generation. As a result, there is a reduced contribution of high performance sequences in the selection probability distribution. The logarithm counteracts this issue by increasing the resolution of the selection probability.

D. Crossover

In each generation $2Q$ offspring are produced from the current population. Members of the population are chosen for reproduction based on their probability of selection. The selection process is constrained such that the crossover procedure only occurs between two distinct members of the population. Members with a high probability of selection not only possess a higher likelihood of reproduction, but also have a higher probability of reproducing with multiple members in a single generation since each crossover is an independent event.

Reproduction is implemented by a crossover between two members in the population, yielding two offspring. To best illustrate the crossover, consider the two chromosomes

$$C_j^{(\alpha)} = \{P_1, \dots, P_i, \dots, P_k\}, \quad (31)$$

$$C_{j'}^{(\alpha)} = \{R_1, \dots, R_i, \dots, R_k\}, \quad (32)$$

where $P_i, R_i \in \mathcal{G}$. The offspring are created by splicing the parent chromosomes at a location chosen at random, where each pulse location has an equal probability of being chosen. Taking the splice point to be the i th pulse site, the resulting offspring are

$$\tilde{C}_j^{(\alpha)} = \{P_1, \dots, P_i, R_{i+1}, \dots, R_k\}, \quad (33)$$

$$\tilde{C}_{j'}^{(\alpha)} = \{R_1, \dots, R_i, P_{i+1}, \dots, P_k\}. \quad (34)$$

It is essential that the offspring still satisfy Eq. (29), however it is not necessarily true that each is guaranteed to do so. If the DD condition is not satisfied, the pulse located at the splice point is manipulated until the condition is satisfied. For example, if $\tilde{C}_j^{(\alpha)}$ does not fulfill the DD condition then it is transformed to

$$\tilde{\tilde{C}}_j^{(\alpha)} = \{P_1, \dots, P'_i, R_{i+1}, \dots, R_k\}, \quad (35)$$

where it now is in agreement with Eq. (29) and $P'_i \in \mathcal{G}$. In the situation that $\tilde{\tilde{C}}_j^{(\alpha)}$ cannot be found, the splice point is chosen again and the process is repeated until the proper offspring are created.

The condition set forth by Eq. (29) restricts the crossover process and in some cases does not allow it at all. By permitting the manipulation of the pulse at the splice point, it is ensured that only the probability of selection dictates reproduction. It is always possible to construct offspring from the above process, since there is no constraint on yielding offspring which are identical to the parent chromosomes. Thus,

every set of parent chromosomes is guaranteed to produce some form of offspring.

Upon producing the $2Q$ offspring, the best $Q/4$ parents and $3Q/4$ offspring are taken to be the new population. The partitioning was chosen based on what appeared to be the most beneficial to the convergence of the algorithm. No duplicate sequences are allowed in the new population, however if the updated population size is less than Q then new members are generated at random from within the reduced search space.

E. Mutation

After reproduction, the new population composed of $Q/4$ parents and $3Q/4$ offspring is used to create $2Q$ mutated sequences, Q single-site and Q double-site. Every sequence in the population participates in both mutation processes, however only a portion of the mutated sequences is retained for the succeeding generation.

Single-site mutations are performed by choosing a pulse site at random and altering the pulse until Eq. (29) is again satisfied. If the DD condition is unsatisfiable then the original pulse is replaced and a different pulse site is chosen. It is possible that only the original configuration satisfies Eq. (29). In this situation the mutated member is simply a duplicate sequence, therefore it is discarded.

Double-site mutations correspond to linked single-site mutations. The process begins in a similar manner by choosing a pulse site at random, say the i th site with pulse P_i . An additional pulse site is now chosen at random from the set of pulse sites which have pulse types equivalent to P_i , e.g., the j th site. Both P_i and P_j are updated simultaneously until the DD condition is again satisfied. If an additional pulse site does not exist, then the initial site is re-selected and the double-site mutation process is repeated. As in the case of the single-site mutation, if the DD condition cannot be satisfied then the mutated sequence is accepted as the original configuration and discarded.

At the conclusion of the two mutations, a portion of the parent, offspring, and mutated sequences will comprise the new population. Only the sequences that have the highest fitness with respect to Eq. (22) are desired from each division of the population. We find that the best $Q/8$ parent, $5Q/8$ offspring, $Q/8$ single-site mutated, and $Q/8$ double-site mutated sequences comprise a favorable distribution for the new population. Other distributions were considered such as taking the best $Q/4$ of all mutated sequences, as well as different proportions of the offspring. However, no distribution appeared to yield a higher probability of optimal sequence convergence.

F. Necessary Convergence Accelerators

As noted above, single-site perturbations may result in large deviations in sequence performance. Hence, the logarithm was introduced to decrease the performance gap between poor- and well-performing sequences, thereby increasing the resolution of the selection probability. However, this

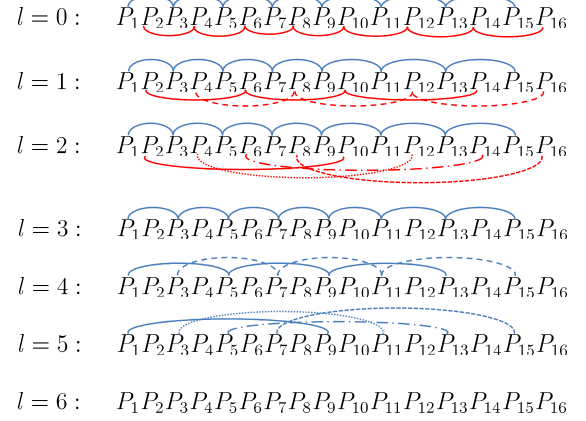


FIG. 1. Complexity-increase protocol for $K = 16$. The upper (blue) curves denote the linking between odd pulse sites and the lower (red) curves denote even pulse site linking. The process begins with all odd and even pulses linked, then continues by removing links between every other even pulse site. Links are removed until all even sites are uncorrelated, thereafter the odd sites undergo the same process.

adjustment only proves to aid in optimal convergence for sequences comprised of $K < 16$ pulses. This is evident from a simple comparison between CDD and numerically located sequences at $K = 16, 64, 256$, where the numerically "optimal" sequences perform far worse than CDD.

We suspect that the local minima convergence is ultimately attributed to significant deviations in sequence performance that result in relatively large local minima traps. We alleviate this complication by introducing two convergence accelerators which act to reduce the size and presence of large traps, thereby smoothening what we refer to as the fitness landscape. Both accelerators are crucial for the algorithm to converge on global optima as the number of pulses increases beyond $K = 16$.

1. Reducing Local Traps via Complexity

Although the size of $\mathcal{N}(K)$ is decreased by imposing the cyclic DD condition, the resulting reduced search space, $\mathcal{N}_R(K)$, maintains its exponential scaling in the number of control pulses. Hence, there is still a high probability of the subspace containing low-performance sequences that lead to large local traps. This issue is resolved by reducing the search space further and systematically increasing its size as the algorithm iterates, such that the search space at the termination of the algorithm is $\mathcal{N}_R(K)$. The additional reduction is achieved by constraining the complexity of the chromosome, thereby moderating the possible sequence configurations.

Initially, we choose each chromosome to represent the most elementary two-dimensional sequence

$$C_j^{(l=0, \alpha=0)} = \{P_{s_1}, P_{s_2}\}, \quad (36)$$

where $l=0$ is the initial complexity index. The notation P_{s_j} denotes a pulse P_j applied at the locations specified by the set s_j . We choose s_1 and s_2 to contain only the odd and even pulse sites, respectively, for the initial population. This is a relevant construction since all known deterministic DD schemes utilizing fixed free intervals contain the same pulse at either every even or odd site [20, 23]. Moreover, it conveniently reduces the space to only $|\mathcal{G}|^2$ sequence configurations for all K .

For a general complexity index l , the chromosome is defined as

$$C_j^{(l,\alpha)} = \{P_{s_1}, P_{s_2}, \dots, P_{s_{\tilde{K}(l)}}\}, \quad (37)$$

where we require that

$$\bigcup_{i=1}^{\tilde{K}(l)} s_i = \text{all sites} \quad \text{and} \quad \bigcap_{i=1}^{\tilde{K}(l)} s_i = \emptyset \quad (38)$$

be satisfied so that only one control pulse is applied at each pulse site. The number of sets $\{s_j\}$ is determined by

$$\tilde{K}(l) = \begin{cases} \frac{3}{2}(l + \frac{4}{3}) : l \text{ even} \\ \frac{3}{2}(l + 1) : l \text{ odd} \end{cases} \quad (39)$$

for $l = 0, 1, 2, \dots, l_{\max}$. At maximum complexity, l_{\max} , the most general sequence within $\mathcal{N}_R(K)$ is permitted. Hence, each s_j is a single element set containing only the j th pulse site. An example of the complexity-increase procedure is illustrated in Figure 1 for $K = 16$. Note that at each level of complexity-increase we have chosen to remove constraints only pertaining to odd or even sites. The constraint between every other even site is removed until each even pulse site is independent, afterwhich the same is performed on the odd sites.

In contrast to Eq. (28), the number of elements in Eq. (37) increases as the algorithm iterates. It is important to note that this aspect does not imply an increase in the number of pulses, rather an increase in the permissible search space. This is an attractive feature since it not only diminishes the presence of local traps, but also yields an initial set of sequence configurations which only scales quadratically in $|\mathcal{G}|$. Choosing sequences for the initial population is obviously much more favorable here since the space is drastically smaller than $\mathcal{N}_R(K)$. In principle, it may even be possible to choose the entire set as the initial population. For a single-qubit system subjected to ideal π -pulses, we find that the complete initial set of configurations is indeed computationally convenient, consisting of only 16 possible configurations. Other pulse profiles lead to larger initial sets, but, remarkably, optimal sequence convergence is possible for initial populations of only 16 sequences.

2. Fitness annealing

Substantial differences in sequence fitness manifest local traps in the fitness landscape. By decreasing the complexity

of the chromosome only the probability of generating local traps is diminished. In order to control the relative differences between high- and low-performance sequences, we introduce an annealing process into the selection probability. Adopted from Ref. [71], the selection probability is redefined as

$$p_j(T) = \frac{\tilde{q}_j(T)}{\sum_i \tilde{q}_i(T)}, \quad (40)$$

[compare with Eq. (30)] such that

$$\tilde{q}_j^{(l,\alpha)}(T) = \exp\left(\frac{q_j^{(\alpha)} - q_{\text{best}}^{(\alpha)}}{T(\alpha)}\right). \quad (41)$$

The performance of the most fit member in the α th generation is denoted by $q_{\text{best}}^{(\alpha)}$ and the temperature function is given by

$$T(\alpha) = T_0 \left(\frac{T_f}{T_0}\right)^{\alpha/\alpha_c} \left[1 - \eta \sin\left(\frac{m\pi}{\alpha_c} \alpha\right)\right]. \quad (42)$$

The temperature function utilized here is a modified version of the one introduced in Ref. [71], where we have included the sinusoidal function to reduce the probability of local minima convergence as $T(\alpha)$ decreases from the initial temperature T_0 to the final temperature T_f . The number of generations between these temperatures is dictated by the cutoff generation α_c , which is chosen based on the value of K . For large K , we pick α_c to be large as well since the annealing process is to accelerate global minimum convergence while reducing the probability of local minimum convergence. The remaining parameters η and m are related to the amplitude and frequency of the oscillations, respectively. Upon increasing the complexity index ℓ , the annealing process resets with an initial temperature T_0 chosen so that all sequences in the current population have an equal likelihood of being chosen for reproduction

IV. OPTIMAL SEQUENCES

In this section, we present numerically optimal π -pulse sequences obtained for the single-qubit system described by Eqs. (15) and (16). First, we consider the case of ideal zero-width pulses. We continue the analysis to include finite-width rectangular profiles, flip-angle errors, and finally the culmination of both types of errors. For all pulse profiles the number of pulses is varied from $K = 1, 2, \dots, 256$ with a pulse-interval $\tau_d = 0.1\text{ns}$.¹

The value of K was varied over a significant range in our simulations, however, we found that only specific values of K are relevant for successive error suppression. In particular, $K_{\text{opt}} = 4, 8, 16, 32, 64, 256$ correspond to the minimum number of pulses required to observe an increase in error suppression or significant improvement in performance. All of

¹ Our choice of units is arbitrary but is meant to be commensurate with electron spin qubits in e.g., quantum dots.

the remaining values of K result in a sequence performance upper bounded by the performance of the previous K_{opt} . For example, the optimal sequences for $K = 17, 18, \dots, 31$ exhibit a performance proportional to that of $K = 16$, if not worse.

The values of K_{opt} were obtained by analyzing the scaling of the performance of the optimal sequences identified at each value of K in the ideal pulse limit. For $K \leq 12$, optimal sequences were located by an exhaustive search, while $K > 12$ demands the use of the GA algorithm discussed above. Upon locating the optimal sequences, the performance measure D is analyzed as a function of τ_d . The order of error suppression, N , is then determined from Eq. (26) by

$$N = \log_{10}(D) - 1. \quad (43)$$

Values of K where N is found to increase ultimately correspond to those identified as values of K_{opt} .

We find that the scaling method described above is a quite convenient numerical method for determining the structure of the dominant term in the effective error Hamiltonian for each of the optimal sequences obtained for a given K_{opt} . In order to fully characterize the scaling of \bar{H}_{err} it is necessary to analyze distance measure D as a function of each relevant parameter. In the case of ideal pulses, this would correspond to analyzing the performance as a function of $\{J, \beta, \tau_d\}$ in the regimes of interaction-dominated dynamics ($J \gg \beta$) and environment-dominated dynamics ($J \ll \beta$) since each regime may exhibit different scalings. When finite pulse duration and flip-angle errors are included the number of parameters increases to either a subset of $\{J, \beta, \tau_d, \tau_p, \epsilon\}$ or the entire set if both forms of pulse errors are present. It is then necessary to analyze the scaling of D in each of the various parameter regimes in addition to the interaction or environment-dominated regimes, e.g., pulse-width dominated ($\tau_p \gg \tau_d$), free evolution dominated ($\tau_p \ll \tau_d$), flip-angle error dominated ($\epsilon \gg J\tau_d$), etc. In the subsequent analysis, we make use of this technique in conjunction with direct calculation of the effective error Hamiltonian to fully characterize sequence performance for each pulse profile. The effective Hamiltonian calculation is utilized to provide additional insight into the structure of \bar{H}_{err} that can not be observed from the scaling method, most notably in situations where multiple sequences exhibit identical performance scalings.

A. Ideal pulses

Here we examine the optimal sequence structure of ideal, zero-width pulses with respect to the strengths of the internal dynamics, J and β , in the regime where $J\tau_d, \beta\tau_d \in [10^{-10}, 10^2]$. Under the condition of uni-axial pulses, the set of possible control pulses $\mathcal{G} = \{I, X, Y, Z\}$, where I is the identity operator and

$$X(Y, Z) = -i\sigma^{x(y,z)} \quad (44)$$

describe π -pulse unitary operators generated by Eq. (17) when $H_C(t)$ is non-zero. Neither the error Hamiltonian, nor pure

environment Hamiltonian, is present during the pulse evolution since the effects of H_0 are negligible in the limit of zero-width (infinite amplitude) pulses.

Characteristics such as the dimension of the reduced search space $\mathcal{N}_R(K)$ are determined by the number of elements in the pulse set \mathcal{G} . Under the conditions of ideal δ -function pulses, $\mathcal{N}_R(K) = 4^{K-1}$ at $l = l_{\text{max}}$. Consequently, the initial search space only contains 16 possible sequence configurations for all K using the complexity-increase procedure described in Sec. III F 1. All 16 are chosen to represent the initial population at the commencement of the algorithm and the size of the population is kept constant throughout.

1. 4-Pulse Sequences

The initial search is performed for $K = 4$ to benchmark the algorithm. The expectation is that the well-known universal decoupling sequence [20]

$$XY_4 = Y f_{\tau_d} X f_{\tau_d} Y f_{\tau_d} X f_{\tau_d} \quad (45)$$

should appear as the optimal sequence since it achieves first order error suppression for general single-qubit decoherence [3]. We confirm that XY_4 is indeed the optimal sequence within the range of J, β given above, along with its obvious generalization

$$GA_4 := P_2 f_{\tau_d} P_1 f_{\tau_d} P_2 f_{\tau_d} P_1 f_{\tau_d}, \quad (46)$$

where $P_1 \neq P_2 \in \{X, Y, Z\}$.

The order of error suppression is confirmed by analyzing the scaling of Eq. (26) with respect to $\{J, \beta, \tau_d\}$. We find

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_d^2) : J \ll \beta \\ \mathcal{O}(J^2\tau_d^2) : J \gg \beta \end{cases}, \quad \text{for } XY_4, \quad (47)$$

These results are consistent with our expectation of first order error suppression and direct calculation of the effective error Hamiltonian, e.g., for $GA_4 = XY_4$:

$$\begin{aligned} \bar{H}_{\text{err}} \approx & -i\tau_d\sigma^x \otimes [B_0, B_x] \\ & + \frac{i}{2}\tau_d\sigma^z \otimes ([B_0, B_z] - i\{B_x, B_y\}). \end{aligned} \quad (48)$$

Additional results are shown in Fig. 2(a), where the performance of GA_4 is shown for a wide range of J, β values.

2. 8-Pulse Sequences

Second order error suppression is achieved for $K = 8$ pulses using the generalized sequence structure

$$GA_8 := P_3(GA_4)P_3(GA_4), \quad (49)$$

such that $P_1, P_2, P_3 \in \{X, Y, Z\}$ and $P_1 \neq P_2$ is implied by the definition of GA_4 . Note that there is no free evolution

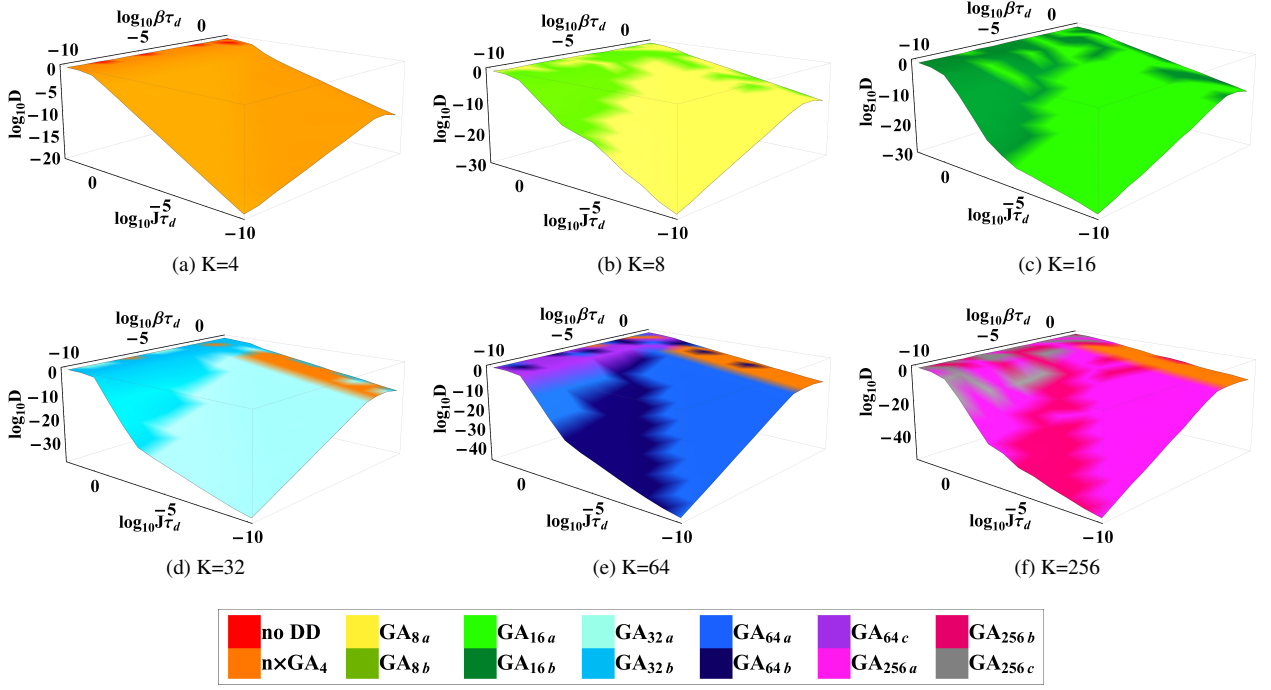


FIG. 2. (color online) Performance of GA_K sequences for $K = 4, 8, 16, 32, 64, 256$, as shown in (a)-(f), respectively, as a function of J and β . The minimum pulse-interval is fixed at $\tau_d = 0.1\text{ns}$ and the results are averaged over 10 realizations of B_μ . The $GA_{K a}$ sequences tend to be optimal in the range $J < \beta$, while $GA_{K b}$ are optimal for $J \geq \beta$. Sequences closely related to the deterministic structure of CDD, $GA_{K c}$, begin to appear as optimal sequences for $J\tau_d > 1$ at $K = 64$. The notation $n \times GA_4$ denotes the application of n cycles of GA_4 and is dependent on the value of K . For the data presented above, $n = 1, 8, 16, 64$ for $K = 4, 32, 64, 256$, respectively.

period between P_3 and P_2 . In Figure 2(b), results for the optimal sequence search are shown with respect to J and β for two versions of Eq. (49):

$$GA_{8a} := IP_1P_2P_1IP_1P_2P_1, \quad (50)$$

when $P_3 = P_2$, and

$$GA_{8b} := (P_3P_2)P_1P_2P_1(P_3P_2)P_1P_2P_1, \quad (51)$$

for $P_3 \neq P_2$. We have dropped the free evolution periods for convenience of notation and highlighted the pulses which are not separated by free evolution periods with parentheses.

The results shown in Figure 2(b) indicate that GA_{8a} is generally optimal when $J\tau_d < 10^{-2}$ and $J < \beta$, while GA_{8b} is optimal for $J\tau_d > 10^{-2}$ and $J > \beta$. The partitioning of the space can be understood in terms of the scaling of the performance, where

$$D \sim \begin{cases} \mathcal{O}(J\beta^2\tau_d^3) & : J \ll \beta \\ \mathcal{O}(J^3\tau_d^3) & : J \gg \beta \end{cases}, \quad \text{for } GA_{8a}, \quad (52a)$$

$$D \sim \mathcal{O}(J\beta^2\tau_d^3) \quad \forall J, \beta, \quad \text{for } GA_{8b}. \quad (52b)$$

We obtain this scaling numerically by analyzing the performance as a function of each of the relevant parameters; see Appendix B for details. As the strength of the error Hamiltonian becomes large, $J\tau_d \gg 1$, it is clearly more beneficial to utilize GA_{8b} since the effective error Hamiltonian scales linearly in J ; as opposed to the cubic scaling of GA_{8a} . Note that

the GA_8 sequences achieve an additional order of suppression relative to XY_4 , as can be seen by comparing the order of τ_d in Eqs. (47) and (52).

Examining the sequence structure of GA_{8a} , a correspondence between this configuration and

$$XY_8 = IXYXIXYX \quad (53)$$

is readily observed when $P_1 = X$ and $P_2 = Y$. Known for providing second order error suppression for general single-qubit decoherence [21], XY_8 gains its decoupling attributes from its structure: the XY_4 sequence followed by a time-reversed copy. An alternative perspective of XY_8 is given by GA_8 , where it appears as a concatenation of XY_4 ($GA_4 = XY_4$) and CPMG ($P_3 = X$) [19]. Depending on the choice of the CPMG pulses two different variations of GA_8 are obtained: GA_{8a} , which is a generalized version of XY_8 , and GA_{8b} . It turns out that concatenation is a key aspect of optimal sequence construction beyond GA_8 and will continue to appear for larger values of K as well.

3. 16-Pulse Sequences

We identify two optimal sequences for $K = 16$: GA_{16a} and GA_{16b} . The former can be expressed in terms of GA_{8a} as

$$GA_{16a} := P_3(GA_{8a})P_3(GA_{8a}), \quad (54)$$

such that $P_3 \neq I$, while the latter sequence is given by

$$GA_{16b} := P_4 (GA_4) P_3 (GA_4) P_4 (GA_4) P_3 (GA_4) \quad (55a)$$

$$= GA_4[GA_4]. \quad (55b)$$

We introduce a useful notation in Eq. (55b) to represent the concatenated structure of Eq. (55a) in a condensed form. The pulse operators for both GA_{16a} and GA_{16b} satisfy $P_j \in \{X, Y, Z\} \forall j$. We point out that the sequences are not required to contain the same pulse operators at each level of concatenation, therefore $P_1 \neq P_3$ and $P_2 \neq P_4$ in Eq. (55b) is also valid.

The regions of optimal performance are depicted in Figure 2(c) for both sequences. Similar to the case of $K = 8$, GA_{16a} is generally optimal for $J < 10^{-2}$ and $\beta > J$, while the remaining portion of the parameter space, $J > 10^{-2}$ and $\beta < J$, is mostly dominated by GA_{16b} . The equivalence between $K = 8$ and $K = 16$ in the J, β parameter space is most likely due to the fact that second order error suppression is still obtained, so the effective error Hamiltonians are quite similar. More specifically, the scaling of the performance for GA_{16a} and GA_{16b} is consistent with GA_{8a} and GA_{8b} , respectively,

$$D \sim \begin{cases} \mathcal{O}(J\beta^2\tau_d^3) & : J \ll \beta \\ \mathcal{O}(J^3\tau_d^3) & : J \gg \beta \end{cases}, \quad \text{for } GA_{16a}, \quad (56a)$$

$$D \sim \mathcal{O}(J\beta^2\tau_d^3) \quad \forall J, \beta, \quad \text{for } GA_{16b}. \quad (56b)$$

Even though an additional order of error suppression is not achieved at $K = 16$ relative to $K = 8$, we include the analysis to show a correspondence between known 16-pulse sequences such as XY_{16} [21] and CDD_2 (see Sec. V), and the GA_{16} sequences. In the definition of Eq. (54) it was noted that $P_3 \neq I$ constitutes a generalized optimal sequence for $K = 16$. Interestingly, the discarded sequence corresponds exactly to XY_{16} . We conclude that XY_{16} is not within the range of optimal sequences due to the fact that it does not reduce errors in any of the three qubit decoherence channels; errors actually accumulate because it is simply two cycles of XY_8 . The second level of concatenation in CDD is found from GA_{16b} when $P_1 = P_3$ and $P_2 = P_4$.

4. 32-Pulse Sequences

A minimum of $K = 32$ pulses is required to achieve third order error suppression. The optimal sequence structures are described by

$$GA_{32a} := GA_4[GA_{8a}] \quad (57)$$

$$GA_{32b} := GA_{8a}[GA_4], \quad (58)$$

where the difference between the two is the order of concatenation. In Fig. 2(d), the regions of optimal performance for GA_{32a} and GA_{32b} indicate a clear distinction between the two sequences and, therefore, concatenation procedure. Optimal performance for GA_{32a} is well-characterized by $J\tau_d < 10^{-3}$ and $\beta\tau_d < 1$. The latter sequence, GA_{32b} , dominates the region where $J\tau_d > 10^{-1} \forall \beta$, and also appears to be optimal in portions of the parameter space where $J\tau_d \geq 10^{-3}$ and

$\beta < J$. The remaining portion of the space, $\beta\tau_d \geq 1$, is dominated by a sequence composed of 8 cycles of GA_4 , denoted by $n \times GA_4$, with $n = 8$, Fig. 2(d). In this particular regime the environment dynamics are fast with respect to the time scale of the pulse delay. The cycle time reduces to effectively combat decoherence on the time scale of the environment dynamics.

The performances for the GA_{32} sequences scale identically, following

$$D \sim \begin{cases} \mathcal{O}(J\beta^3\tau_d^4) & : J \ll \beta \\ \mathcal{O}(J^2\beta^2\tau_d^4) & : J \gg \beta. \end{cases} \quad (59)$$

for $J\tau_d < 1$ and $\beta\tau_d < 1$. Differences in the performance of equal order sequences can be due to differences in the “performance amplitude,” namely, the fact that the effective error Hamiltonian can contain the same error terms with larger coefficients, hence reduced performance. Furthermore, higher order contributions can play a role in performance variations as $\beta\tau_d \rightarrow 1$ and $J\tau_d \rightarrow 1$.

5. 64-Pulse Sequences

Optimal sequences located at $K = 64$ acquire fourth order error suppression via the generalized sequence structures

$$GA_{64a} := GA_{8a}[GA_{8a}], \quad (60)$$

$$GA_{64b} := GA_{8b}[GA_{8b}], \quad (61)$$

each of which correspond to a first level concatenation of GA_{8a} or GA_{8b} , respectively. Additional combinations of concatenation, e.g., $GA_{8a}[GA_{8b}]$ and $GA_{8b}[GA_{8a}]$ turn out not to be optimal since only third order error suppression is achieved. The pulse operators satisfy $P_j \in \{X, Y, Z\} \forall j$ as in previous cases.

Sequence performance for both GA_{64a} and GA_{64b} for $J\tau_d \ll 1$ and $\beta\tau_d \ll 1$ scales as

$$D \sim \begin{cases} \mathcal{O}(J\beta^4\tau_d^5) & : J \ll \beta \\ \mathcal{O}(J^3\beta^2\tau_d^5) & : J \gg \beta, \end{cases} \quad (62)$$

where GA_{64a} is optimal in the region of $J \leq \beta$ for $\beta\tau_d < 1$ and GA_{64b} is optimal for $J > \beta$ for $J\tau_d < 1$. The distinction between the regions of optimal performance, shown in Fig. 2(e), is due to a transition in performance at $J = \beta$ that results in GA_{64b} becoming the preferred construction. We suspect that the effective Hamiltonian for GA_{64a} contains terms which result in a J^2 -scaling in D , in addition to the dominant J^3 term, that diminishes performance once in the $J > \beta$ regime. The GA_{64b} construction does not receive a significant contribution from the quadratic term, if any at all; hence, the increase in performance.

There are two additional sequences shown in Fig. 2(e). The $n \times GA_4$ sequence denotes the application of $n = 16$ cycles of GA_4 to obtain the 64 pulses required for the search. This particular sequence construction is optimal for $\beta\tau_d > 1 \forall J$, as in the case of $K = 32$. The remaining sequence

$$GA_{64c} = GA_4[GA_4[GA_4]] \quad (63)$$

obtains third order error suppression and exhibits a performance scaling as

$$D \sim \begin{cases} \mathcal{O}(J\beta^3\tau_d^4) & : J \ll \beta \\ \mathcal{O}(J^2\beta^2\tau_d^4) & : J \gg \beta \end{cases} \quad (64)$$

The region of optimal performance is well-characterized by $J\tau_d > 1 \forall \beta$, the high-frequency coupling regime. Note the correspondence between GA_{64c} and the third level of concatenation in CDD. Similar to $K = 16$, we find that CDD remains an optimal sequence construction in specific parameter regimes.

6. 256-Pulse Sequences

The final optimal sequence structures we have obtained are for $K = 256$. All sequences are classified into three general structures denoted by GA_{256a} , GA_{256b} , GA_{256c} . The GA_{256a} and GA_{256b} constructions achieve fifth order error suppression by concatenating GA_{64a} with GA_4 as

$$\begin{aligned} GA_{256a} &= GA_4[GA_{8a}[GA_{8a}]] \\ &= GA_4[GA_{64a}] \end{aligned} \quad (65)$$

and concatenating GA_{32b} with GA_{8b} to obtain

$$\begin{aligned} GA_{256b} &= GA_{8b}[GA_{8a}[GA_4]] \\ &= GA_{8b}[GA_{32a}]. \end{aligned} \quad (66)$$

The remaining sequence structure,

$$GA_{256c} = GA_4[GA_4[GA_4[GA_4]]], \quad (67)$$

attains fourth order error suppression and is essentially a generalized version of the fourth level of concatenation of CDD. As in $K = 64$, $P_j \in \{X, Y, Z\} \forall j$.

In examining the performance as a function of J and τ_d for GA_{256a} , we find the scaling

$$D \sim \begin{cases} \mathcal{O}(J\beta^5\tau_d^6) & : J \ll \beta \\ \mathcal{O}(J^3\beta^3\tau_d^6) & : J \gg \beta, \end{cases} \quad (68)$$

and a similar scaling for the performance for GA_{256b} up to $J\tau_d > 10^{-4}$. The performance for GA_{256c} scales as

$$D \sim \begin{cases} \mathcal{O}(J\beta^4\tau_d^5) & : J \ll \beta \\ \mathcal{O}(J^2\beta^3\tau_d^5) & : J \gg \beta \end{cases} \quad (69)$$

As found for $K = 64$, successive levels of GA_4 concatenation, i.e., GA_{256c} , is predominantly optimal when $J\tau_d \geq 1$. Note that in the region where $\beta\tau_d > 1$, $64 \times GA_4$ (64 cycles) remains the optimal sequence $\forall J$.

The regions of optimal performance are depicted in Fig. 2(f). Two transitions in sequence performance are observed: one at $J = \beta$, where GA_{256b} becomes the preferred sequence structure, and a second at $J\tau = 10^{-4}$, where GA_{256a} regains dominance and the performance of GA_{256b} diminishing rapidly. Although the performance scales equivalently for GA_{256a} and GA_{256b} , there is an obvious inherent

Sequence	$J \ll \beta$	$J \gg \beta$
GA_4	$\mathcal{O}(J\beta\tau_d^2)$	$\mathcal{O}(J^2\tau_d^2)$
GA_{8a}	$\mathcal{O}(J\beta^2\tau_d^3)$	$\mathcal{O}(J^3\tau_d^3)$
GA_{8b}	$\mathcal{O}(J\beta^2\tau_d^3)$	$\mathcal{O}(J\beta^2\tau_d^3)$
GA_{16a}	$\mathcal{O}(J\beta^2\tau_d^3)$	$\mathcal{O}(J^3\tau_d^3)$
GA_{16b}	$\mathcal{O}(J\beta^2\tau_d^3)$	$\mathcal{O}(J\beta^2\tau_d^3)$
GA_{32a}	$\mathcal{O}(J\beta^3\tau_d^4)$	$\mathcal{O}(J^2\beta^2\tau_d^4)$
GA_{32b}	$\mathcal{O}(J\beta^3\tau_d^4)$	$\mathcal{O}(J^2\beta^2\tau_d^4)$
GA_{64a}	$\mathcal{O}(J\beta^4\tau_d^5)$	$\mathcal{O}(J^3\beta^2\tau_d^5)$
GA_{64b}	$\mathcal{O}(J\beta^4\tau_d^5)$	$\mathcal{O}(J^3\beta^2\tau_d^5)$
GA_{64c}	$\mathcal{O}(J\beta^3\tau_d^4)$	$\mathcal{O}(J^2\beta^2\tau_d^4)$
GA_{256a}	$\mathcal{O}(J\beta^5\tau_d^6)$	$\mathcal{O}(J^3\beta^3\tau_d^6)$
GA_{256b}	$\mathcal{O}(J\beta^5\tau_d^6)$	$\mathcal{O}(J^3\beta^3\tau_d^6)$
GA_{256c}	$\mathcal{O}(J\beta^4\tau_d^5)$	$\mathcal{O}(J^2\beta^3\tau_d^5)$

TABLE I. Summary of distance measure (D) scalings for each optimal GA_K sequence identified in the ideal pulse limit, for a fixed pulse-interval of τ_d .

weakness in GA_{32a} since further concatenation reduces the region of optimal performance; compare Figs. 2(d) and (f). The placement of GA_4 in the concatenation is believed to ultimately attribute to the attenuated performance of GA_{256b} . We suspect that there will be an eventual critical level of concatenation after which optimality no longer exists for concatenated versions of GA_{32a} due to the first order error suppression of GA_4 . (This effect is similar to that seen for asymmetric QDD, where the performance is bounded by the order of error suppressing corresponding to the nested sequence with the smallest number of pulses [40].)

We summarize the scaling of D for each of the optimal sequences discussed above for all K_{opt} in Table I. The optimal sequences located here achieve an additional order of decoupling, on average, over the standard XY_4 -based CDD sequences. For example, compare GA_{64c} and GA_{256c} to the additional optimal sequences.

7. Interlude: optimal CDD sequences

Having found that the shortest sequences achieving order $N=1, 2, 3, 4, 5$ decoupling have length $K = 4, 8, 32, 64, 256$, respectively, we now consider what is the optimal sequence length for concatenation at fixed pulse-interval τ_d . A base $DD(K, N)$ sequence of length K achieving N th order decoupling uses K^q pulses (and hence total time $K^q\tau_d$) to achieve order $N \times q$ decoupling at concatenation level q , a sequence we denote $CDD_q(K, N)$.

Let us first compare $K = 4, 8$ at their lowest common equal decoupling order, i.e., at order 2. At this order the $K = 4$ sequence is $CDD_2(4, 1)$ and requires $4^2 = 16$ pulses, and the $K = 8$ sequence is $CDD_1(8, 2)$ and requires 8 pulses. Thus the $K = 8$ sequence is optimal: the comparison at equal higher order decoupling levels 2ℓ , $\ell > 1$, yields $16^{2\ell} = 2^{8\ell}$ pulses for $CDD_{2\ell}(4, 1)$, compared to $8^{2\ell} = 2^{6\ell}$ for $CDD_\ell(8, 2)$.

Decoupling Order	Number of Pulses				
	GA_4	GA_8	GA_{32}	GA_{64}	GA_{256}
1	$4^1 = 2^2$				
2	$4^2 = 2^4$	$8^1 = 2^3$			
3	$4^3 = 2^6$		$32^1 = 2^5$		
4	$4^4 = 2^8$	$8^2 = 2^6$		$64^1 = 2^6$	
5	$4^5 = 2^{10}$				$256^1 = 2^8$
6	$4^6 = 2^{12}$	$8^3 = 2^9$	$32^2 = 2^{10}$		
8	$4^8 = 2^{16}$	$8^4 = 2^{12}$		$64^2 = 2^{12}$	
10	$4^{10} = 2^{20}$	$8^5 = 2^{15}$			$256^2 = 2^{16}$
12	$4^{12} = 2^{24}$	$8^6 = 2^{18}$	$32^4 = 2^{20}$	$64^3 = 2^{18}$	
15	$4^{15} = 2^{30}$		$32^5 = 2^{25}$		$256^3 = 2^{24}$
20	$4^{20} = 2^{40}$	$8^{10} = 2^{30}$		$64^5 = 2^{30}$	$256^4 = 2^{32}$
24	$4^{24} = 2^{48}$	$8^{12} = 2^{36}$	$32^8 = 2^{40}$	$64^6 = 2^{36}$	
30	$4^{30} = 2^{60}$	$8^{15} = 2^{45}$	$32^{10} = 2^{50}$		$256^6 = 2^{48}$
40	$4^{40} = 2^{80}$	$8^{20} = 2^{60}$		$64^{10} = 2^{60}$	$256^8 = 2^{64}$
45	$4^{45} = 2^{90}$		$32^{15} = 2^{75}$		$256^9 = 2^{72}$
60	$4^{60} = 2^{120}$	$8^{30} = 2^{90}$	$32^{20} = 2^{100}$	$64^{15} = 2^{90}$	$256^{12} = 2^{96}$

TABLE II. Summary of decoupling efficiencies for ideal-pulse concatenated sequences. The first column gives the achievable decoupling order, the next five columns give the number of pulses required by each of the GA_K base sequences to achieve the desired decoupling order. The exponent q in each entry of the form K^q is the concatenation level. The shortest sequence lengths needed for a given decoupling order are boxed.

When we include DD(32, 3) the lowest common decoupling order is increased to 6. At this order the $K = 4$ sequence is $CDD_6(4, 1)$ and requires $4^6 = 2^{12}$ pulses, the $K = 8$ sequence is $CDD_3(8, 2)$ and requires $8^3 = 2^9$ pulses, and the $K = 32$ sequence is $CDD_2(32, 3)$ and requires $32^2 = 2^{10}$ pulses. Thus the $K = 8$ sequence is again optimal.

When we include DD(64, 4) the lowest common decoupling order is increased to 12. At this order the $K = 4$ sequence is $CDD_{12}(4, 1)$ and requires $4^{12} = 2^{24}$ pulses, the $K = 8$ sequence is $CDD_6(8, 2)$ and requires $8^6 = 2^{18}$ pulses, the $K = 32$ sequence is $CDD_4(32, 3)$ and requires $32^4 = 2^{20}$ pulses, and the $K = 64$ sequence is $CDD_3(64, 4)$ and requires $64^3 = 2^{18}$ pulses. Thus the $K = 8$ and $K = 64$ sequences are optimal.

Finally, when we include DD(256, 5) the lowest common decoupling order is increased to 60. At this order the $K = 4$ sequence is $CDD_{60}(4, 1)$ and requires $4^{60} = 2^{120}$ pulses, the $K = 8$ sequence is $CDD_{30}(8, 2)$ and requires $8^{30} = 2^{90}$ pulses, the $K = 32$ sequence is $CDD_{20}(32, 3)$ and requires $32^{20} = 2^{100}$ pulses, the $K = 64$ sequence is $CDD_{15}(64, 4)$ and requires $64^{15} = 2^{90}$ pulses, and the $K = 256$ sequence is $CDD_{12}(256, 5)$ and requires $256^{12} = 2^{96}$ pulses. Thus the $K = 8$ and $K = 64$ sequences are once again optimal.

An exhaustive list of all lowest common decoupling orders generated from the base orders $N \in \{1, 2, 3, 4, 5\}$ is $M \in \{1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 30, 40, 45, 60\}$. It is straightforward to check that whenever $CDD_q(8, 2)$ achieves one of these decoupling orders, i.e., whenever M is even, then $CDD_q(8, 2)$ requires the smallest number of pulses in

comparison to all other $CDD_q(K, N)$ sequences capable of achieving the same decoupling order (i.e., with K dividing M). The same is true for $CDD_q(64, 4)$, which requires the same number of pulses as $CDD_q(8, 2)$, except that it cannot be used when $M \in \{10, 30\}$ (not divisible by 4). The only exceptions to the superiority of $CDD_q(8, 2)$ [or $CDD_q(64, 4)$] are, of course, the cases with M odd, namely $M = 3$ and $M \in \{5, 15\}$, where the optimal sequences are, respectively, $CDD_q(32, 3)$ and $CDD_q(256, 5)$. A complete classification is presented in Table II.

B. Finite-width Pulses

We now extend the analysis to include errors due to finite-width rectangular pulses of duration τ_p . We note that in this case EDD is a known way to achieve first order pulse-width error suppression, with the added assumption that pulse shaping is possible [45]. Each optimal sequence construction is examined with respect to $J/\beta \in [10^{-15}, 10^3]$ for $\beta = 1$ kHz and $\tau_p/\tau_d \in [10^{-6}, 10^3]$ for $\tau_d = 0.1$ ns. The set of allowable control pulses

$$\mathcal{G} = \{I, X, Y, Z, \bar{X}, \bar{Y}, \bar{Z}\} \quad (70)$$

contains three additional pulses $\{\bar{X}, \bar{Y}, \bar{Z}\}$ corresponding to 180-degree phase flips of the non-trivial pulses $\{X, Y, Z\}$, which are generated by taking $A \rightarrow -A$ in Eq. (19). Note that in contrast to the ideal pulse case, the identity operator $I \equiv e^{-i\tau_p H_0}$ so that for a given K the cycle time is equivalent

for all sequences. The remaining unitary pulse operators are generated from Eq. (19) and given by

$$X(Y, Z) = e^{-i\tau_p(A\sigma^{x(y,z)}+H_0)}, \quad (71)$$

where the internal Hamiltonian H_0 is now included due to the fact that the pulses are now finite in amplitude. This form can be reconciled with the ideal pulse operators given in Eq. (44) by considering infinite pulse amplitude ($A \rightarrow \infty$) and a pulse duration that is effectively zero ($\tau_p \rightarrow 0$) so that $\|H_C(t)\|_{\tau_p} = \pi/2$ is maintained and $\|H_0\|_{\tau_p} \rightarrow 0$.

The finite amplitude assumption essentially defines a pulse duration that differs from zero, so that H_0 becomes a relevant contributor to the dynamics generated by the pulse. Of course we must still make the strong pulse assumption $\|H_C(t)\| \gg \|H_0\|$ in order to reduce the additional errors generated by H_0 during the pulse. We enforce this assumption in the following section and utilize it to calculate effective pulse dynamics where the contributions of H_0 are essentially perturbations to $H_C(t)$. The effective pulse operators are then used to calculate effective error Hamiltonians for each finite-width-robust GA sequence.

The dimension of the reduced search space is increased by the additional pulses to $\mathcal{N}_R(K) = 7^{K-1}$ at l_{\max} . Consequently, the initial search space contains $7^2 = 49$ configurations at all values of K . We find that not all 49 sequences are required to obtain optimal sequences. Surprisingly, an initial population of 16 sequences chosen at random from the 49 is sufficient to converge on global extrema for all K .

1. 4-Pulse Sequences

The optimal sequence structures for $K = 4$ is nearly equivalent to the ideal pulse case, differing by phase adjustments to effectively combat the additional error introduced by pulse width. The generalized structures are given by

$$RGA_4 := \bar{P}_2 P_1 \bar{P}_2 P_1, \quad (72)$$

$$RGA_{4'} := \bar{P}_2 \bar{P}_1 \bar{P}_2 P_1, \quad (73)$$

where we refer to the pulse error-optimized sequences as robust GA_K (RGA_K). In order to determine the order of error suppression associated with the pulse-width, we examine the scaling of the performance as a function of τ_p . The performance is found to scale according to

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_d^2) & : J \ll \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\tau_p) & : J \ll \beta, \tau_p \gg \tau_d \\ \mathcal{O}(J^2\tau_d^2) & : J \gg \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\tau_p) & : J \gg \beta, \tau_p \gg \tau_d \end{cases} \quad (74)$$

for both sequences, which is in agreement with direct calculation of the effective error Hamiltonian for each sequence when $\{P_1, P_2\} = \{X, Y\}$:

$$\bar{H}_{\text{err}}^{RGA_4} \approx \frac{4\tau_p}{\pi\tau_c} \sigma^z B_{x-y} + \bar{H}_{\text{err}}^{GA_4} \quad (75)$$

$$\bar{H}_{\text{err}}^{RGA_{4'}} \approx \frac{4\tau_p}{\pi\tau_c} (\sigma^y B_z + \sigma^z B_x) + \bar{H}_{\text{err}}^{GA_4}, \quad (76)$$

where $\tau_c = 4(\tau_p + \tau_d)$ and $B_{x-y} \equiv B_x - B_y$. A distinction between the two sequences is apparent at first order in τ_p , where the dominant error term for RGA_4 , isolated along the σ^z error channel, is a difference between two bath operators and the additional phase adjustment included in $RGA_{4'}$ generates a sum of errors along two error channels: σ^y and σ^z . The former is preferable since the difference bath operator $B_x - B_y$ can vanish, while the sum operator appearing in the expansion of $RGA_{4'}$ cannot. For randomly selected bath operators this translates into a smaller operator norm, on average. The resulting difference between the two sequences is indeed very visible in Fig. 3(a), which shows the regions of optimal performance. $RGA_{4'}$ is generally optimal for $J\tau_d \sim 1$ and $J \gg \beta$; the remaining region is dominated by RGA_4 .

2. 8-Pulse Sequences

Three optimal sequences are identified for $K = 8$ which possess similar characteristics to GA_{8a} , GA_{8b} , and a well-known 8-pulse decoupling sequence designed to combat finite-width pulses. Adjusting pulse phases to obtain robustness against finite-width pulse errors, the generalized sequence structures are

$$RGA_{8a'} := IP_1 \bar{P}_2 P_1 IP_1 \bar{P}_2 P_1 \quad (77)$$

$$RGA_{8b} := \bar{P}_3 (RGA_4) P_3 (RGA_4), \quad (78)$$

$$RGA_{8c} := P_1 P_2 P_1 P_2 P_2 P_1 P_2 P_1 \quad (79)$$

The first sequence identified is a robust version of GA_{8a} , differing from the ideal pulse construction by the two inverted P_2 pulses. The prime is used in anticipation of later results where RGA_{8a} is introduced in Eq. (105) below. The RGA_{8b} sequence utilizes a concatenation between a robust version of generalized CPMG, $\bar{P}_3 f_{\tau_d} P_3 f_{\tau_d}$, to encompass RGA_4 , while RGA_{8c} denotes a generic version of the Eulerian DD (EDD) sequence derived using walks on Cayley graphs [45].

$RGA_{8a'}$ performance scales according to

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_p\tau_d) & : J \ll \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\beta\tau_p^2) & : J \ll \beta, \tau_p \gg \tau_d \\ \mathcal{O}(J^2\tau_p\tau_d) & : J \gg \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J^2\tau_p^2) & : J \gg \beta, \tau_p \gg \tau_d \end{cases}, \quad (80)$$

which is dictated by the effective error Hamiltonian

$$H_{\text{err}}^{RGA_{8a'}} \approx \frac{16\tau_d\tau_p}{\pi\tau_c} \sigma^y B_x^2 - \frac{8\tau_d\tau_p}{\pi\tau_c} \sigma^y \{B_x, B_y\} + \mathcal{O}(J\beta\tau_d\tau_p) \quad (81)$$

computed for the specific case of $P_1 = X$ and $P_2 = Y$. First and second order error suppression in τ_p and τ_d is achieved, respectively, yielding the quadratic term of $\mathcal{O}(J^2\tau_d\tau_p)$ as the dominant decoherence generator when $J \gg \beta$ in the strong pulse regime ($\tau_p \ll \tau_d$). As we will illustrate below, $RGA_{8a'}$ is the preferred sequence construction for a majority of the parameter regimes considered due to the fact that it provides considerable suppression of $\mathcal{O}(J^2\tau_d\tau_p)$ terms over RGA_{8b} and RGA_{8c} .

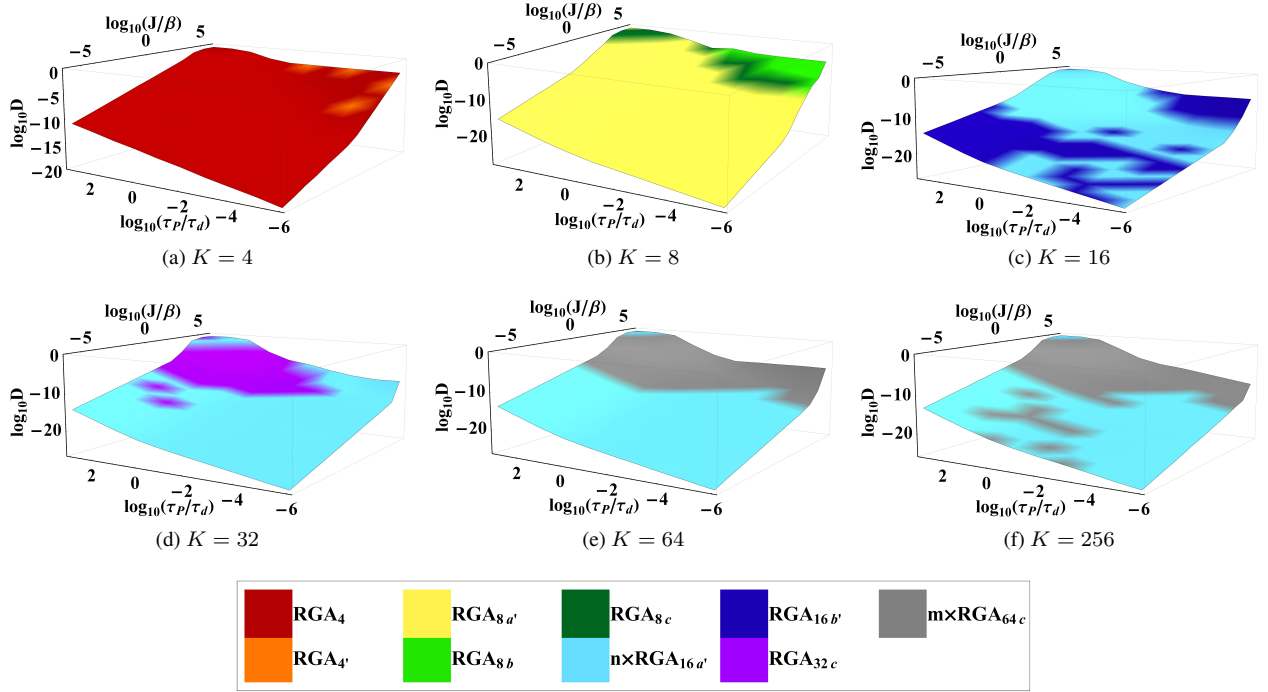


FIG. 3. (color online) Performance of optimal RGA_K sequences for $K = 4, 8, 16, 32, 64, 256$ shown in (a)-(f), respectively, as a function of J/β and τ_p/τ_d when DD is subjected to finite pulse duration. The norm of the bath Hamiltonian is fixed at $\beta = 1\text{kHz}$, while $J/\beta \in [10^{-6}, 10^6]$. The pulse-interval $\tau_d = 0.1\text{ns}$ and the pulse width is varied in the range $\tau_p/\tau_d \in [10^3, 10^{-6}]$. For a given K , the optimal sequence configuration is most sensitively dependent upon variations in J . Contrary to the ideal pulse case, concatenated structures composed of $RGA_{8a'}$ and RGA_{8c} appear to be the most favorable, in particular for $K \geq 16$ where $RGA_{16a'}$ and RGA_{64c} repeatedly emerge as optimal sequences. Sequence performance saturates at $K = 16$, while robustness begins to diminish at $K = 256$. $n \times RGA_{16b'}$ denotes n cycles of $RGA_{16b'}$; $n = 1, 4, 16$ for $K = 16, 64, 256$ respectively. The notation is similar for $m \times RGA_{64c}$, where $m = 1, 4$ cycles are used for $K = 64, 256$.

For the second sequence, RGA_{8b} , the performance is dictated by the distance measure scaling

$$D \sim \mathcal{O}(J\tau_p) \quad \forall J, \beta, \tau_p, \tau_d, \quad (82)$$

where the elementary concatenation procedure appears to only benefit the suppression of terms solely proportional to the pulse-interval. Computing the effective error Hamiltonian for RGA_{8b} for both possible sequence variations, $P_3 = P_1$ and $P_3 \neq P_1$, we find

$$H_{\text{err}, P_3=X}^{RGA_{8b}} \approx \frac{4\tau_p}{\pi\tau_c} (\sigma^x B_z + \sigma^y B_x) \quad (83a)$$

$$H_{\text{err}, P_3=Z}^{RGA_{8b}} \approx \frac{4\tau_p}{\pi\tau_c} [\sigma^y B_z + \sigma^z (B_x - 2B_y)] \quad (83b)$$

for the specific case of $\{P_1, P_2\} = \{X, Y\}$. Note that by definition $P_3 \neq P_2$ for RGA_{8b} . Adhering to the scaling demonstrated numerically by Eq. (82), both sequence variations convey a linear scaling in τ_p that is accompanied by a sum of two decoherence-generating terms which are distributed among two error channels. The primary distinction between Eqs. (83a) and (83b) is the difference bath operator $B_x - 2B_y$, which appears in the $P_3 = Z$ case. On average, we can expect this term to exhibit a larger operator norm and, therefore, a less desirable performance than the $P_3 = X$ case.

Although such differences in sequence design can be expected to effect the structure of the effective error Hamiltonian, it is seemingly unclear how one would determine which configuration is the most favorable a priori. In the discussions on flip-angle errors, we address this issue again and show that such adjustments may, in fact, result in very different distance measure scalings.

The remaining sequence, RGA_{8c} , attains first order error suppression in τ_p by traversing the Cayley graph $\Gamma = \Gamma(\mathcal{S}, \mathcal{G})$, where \mathcal{S} denotes the single-qubit Pauli group with elements denoting the graph vertices and $\mathcal{G} = \{I, X, Y, Z\}$ is the generating set comprising the edges. The original construction (captured by RGA_{8c}) utilized two closed Eulerian cycles on Γ such that the second is completed by returning along the first path. However, additional paths exist which do not require closed cycles to obtain first order suppression in τ_p . One such case is $RGA_{8a'}$, where the initial path and its inversion are both *open* Eulerian paths, as illustrated in Fig. 4. In comparing $RGA_{8a'}$ and RGA_{8c} , the most important aspect appears to be the manner in which the paths are traversed rather than their closure. Note that variations in pulse phases may aid in the error suppression process, but are not necessarily required to obtain first order decoupling in τ_p as $G A_{8a}$ also performs the task.

In terms of performance scaling, RGA_{8c} results in

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_d^2) : J \ll \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\beta\tau_p^2) : J \ll \beta, \tau_p \gg \tau_d \\ \mathcal{O}(J^2\tau_d^2) : J \gg \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J^2\tau_p^2) : J \gg \beta, \tau_p \gg \tau_d \end{cases}, \quad (84)$$

where first order decoupling in τ_p and τ_d is clearly indicated. Note that second order error suppression in τ_d is not obtained for RGA_{8c} , hence the advantage of $RGA_{8a'}$ in the strong pulse regime. As $\tau_p \rightarrow \tau_d$ when $J \gg \beta$ the effective error Hamiltonian for RGA_{8c} is essentially given by

$$H_{\text{err}}^{RGA_{8c}} \approx -\frac{4\tau_d\tau_p}{\pi\tau_c}\sigma^x(B_y^2 + \{B_x, B_y\}) - \frac{4\tau_d\tau_p}{\pi\tau_c}\sigma^y(B_x^2 + \{B_x, B_y\}) + \mathcal{O}(J^2\tau_d^2) \quad (85)$$

Note the larger number of effective system-bath interaction terms present for RGA_{8c} as compared to the effective error Hamiltonian of $RGA_{8a'}$, described in Eq. (81). A similar analysis can also be done in the $\tau_p \gg \tau_d$ regime to further convey the superiority of $RGA_{8a'}$ over RGA_{8b} and RGA_{8c} .

In Fig. 3(b), we support the above results for $RGA_{8a'}$, RGA_{8b} and RGA_{8c} with a numerical analysis over a wide range of J and τ_p . As suspected from the performance scalings [Eqs. (80)-(84)] and effective error Hamiltonians [Eqs. (81)-(85)], $RGA_{8a'}$ is the preferred sequence for a majority of the parameter space. RGA_{8b} and RGA_{8c} emerge as the optimal sequence configurations as $J \gg \beta$, with RGA_{8b} dominating the $\tau_p \ll \tau_d$ regime and the first order suppression of RGA_{8c} in τ_p becoming more favorable when $\tau_p \gg \tau_d$.

3. 16-Pulse Sequences

Robust sequences at $K = 16$ can be expressed in terms of $RGA_{8a'}$ and first level concatenations of RGA_4 , and $RGA_{4'}$. The generalized sequence structures are given by

$$RGA_{16a'} := P_3(RGA_{8a'})P_3(RGA_{8a'}) \quad (86)$$

$$RGA_{16b'} := RGA_4[RGA_{4'}] \quad (87)$$

where the primes denotes the difference in phase configuration among the pulses as described in Sec. IV B 1 and IV B 2. The first sequence, $RGA_{16a'}$, is a concatenation between $P_3 f_{\tau_d} P_3 f_{\tau_d}$ and $RGA_{8a'}$, while $RGA_{16b'}$ is a single-level concatenation that involves both RGA_4 and $RGA_{4'}$. This particular configuration is found to be more suitable for combating pulse-width errors since it allows for first order suppression of terms proportional to τ_p while additional configurations, such as $RGA_{4'}[RGA_{4'}]$ and $RGA_{4'}[RGA_4]$, do not. Note that $RGA_4[RGA_4]$ does not appear either even though it achieves first order error suppression in τ_p due to the more significant error suppression provided by both $RGA_{16a'}$ and $RGA_{16b'}$.

As shown in Fig. 3(c), the regions of optimal performance are essentially divided equally among $RGA_{16a'}$ and $RGA_{16b'}$. $RGA_{16a'}$ tends to be optimal for $J \geq \beta \forall \tau_p$, while

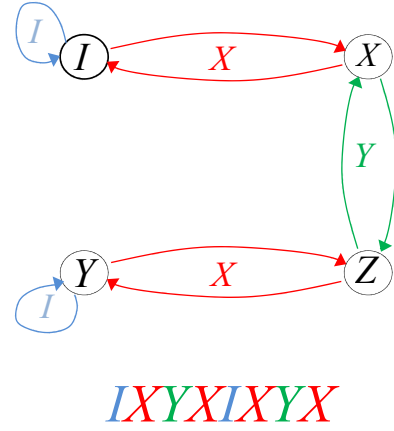


FIG. 4. (color online) Pictorial depiction for the action of $RGA_{8a'}$ as an Eulerian path along the Cayley graph $\Gamma(\mathcal{S}, \mathcal{G})$ with vertices $\mathcal{S} = \{I, X, Y, Z\}$ and generating set $\mathcal{G} = \{I, X, Y, Z\}$. Note that unlike the EDD construction [Eq. (79)], $RGA_{8a'}$ is generated by Eulerian paths, rather than cycles.

$RGA_{16b'}$ dominates the region $J < \beta$. It is not possible to deduce an explanation for the partitioning of the space from the numerical scaling method since both sequences exhibit the same distance measure scaling

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_p\tau_d) : J \ll \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\beta\tau_p^2) : J \ll \beta, \tau_p \gg \tau_d \\ \mathcal{O}(J\beta\tau_p\tau_d) : J \gg \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J^2\tau_p^2) : J \gg \beta, \tau_p \gg \tau_d \end{cases}. \quad (88)$$

In order to gain more insight, we turn to the effective error Hamiltonian for each sequence and determine that

$$\begin{aligned} \bar{H}_{\text{err}}^{RGA_{16a'}} \approx & \frac{8i\tau_d\tau_p}{\tau_c} \left(\sigma^x[B_0, B_x - \frac{2}{\pi}B_z] \right. \\ & \left. + \sigma^y[B_0, B_y - \frac{2}{\pi}B_z] + \sigma^z[B_0, B_z + \frac{2}{\pi}B_y] \right) \end{aligned} \quad (89)$$

and

$$\begin{aligned} \bar{H}_{\text{err}}^{RGA_{16b'}} \approx & \frac{8i\tau_p\tau_d}{\tau_c} \sigma^x[B_0, B_x - \frac{2}{\pi}B_y - \frac{4}{\pi}B_z] \\ & + \frac{8i\tau_p\tau_d}{\tau_c} \sigma^y[B_0, B_y - \frac{2}{\pi}B_x] \\ & + \frac{8i\tau_p\tau_d}{\tau_c} \sigma^z[B_0, B_z - \frac{2}{\pi}B_x]. \end{aligned} \quad (90)$$

in the strong-pulse regime. The primary distinction between the two constructions is the additional term in the σ^x -channel for $RGA_{16b'}$, which explains the preference in $RGA_{16a'}$ when $J > \beta$ and the relatively equal sharing of the parameter space when $J < \beta$. In the remaining region where $\tau_p > \tau_d$ partitioning of the space can also be explained in terms of the

effective error Hamiltonian, where

$$\bar{H}_{\text{err}}^{RGA_{16a'}} \sim \frac{\tau_p^2}{\tau_c} \sigma^x (B_y^2 + B_z^2) \quad (91)$$

$$\bar{H}_{\text{err}}^{RGA_{16b'}} \sim \frac{\tau_p^2}{\tau_c} \sigma^z (B_x^2 + B_y^2 + B_z^2 + \{B_y, B_z\}) \quad (92)$$

clearly agrees with the more pronounced dominance of $RGA_{16a'}$ when $J > \beta$ indicated in the numerical results. An additional analysis confirms a higher degree of error suppression for $RGA_{16b'}$ when $J < \beta$ and $\tau_p > \tau_d$.

The suppression of $\mathcal{O}(J^2 \tau_d \tau_p)$ terms is the most substantial benefit of the $K = 16$ sequences over $K = 8$. $RGA_{16a'}$ achieves this additional decoupling by concatenating $RGA_{8a'}$ with an elementary one-dimensional sequence to suppress the single-channel decoherence of $\mathcal{O}(J^2 \tau_d \tau_p)$ remaining in Eq. (81). $RGA_{16b'}$ requires a more complicated concatenation process to attain a similar effective error Hamiltonian since it is also necessary to remove the $\mathcal{O}(J \tau_p)$ terms not suppressed by $RGA_{4'}$.

4. 32-Pulse Sequences

Optimal performance at $K = 32$ is characterized by two particular sequence constructions: $2 \times RGA_{16a'}$ and

$$RGA_{32c} = RGA_{8c}[RGA_4]. \quad (93)$$

As shown in Fig. 3(d), $RGA_{16a'}$ is the preferred choice for $J < \beta \forall \tau_p, \tau_d$, as well as $J > \beta$ when $\tau_p \ll \tau_d$, while RGA_{32c} dominates the $J > \beta$ with $\tau_p > \tau_d$ regime. Using the scaling of $RGA_{16a'}$ given in Eq. (88) and

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_d\tau_p) & : J \ll \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\beta\tau_p^2) & : J \ll \beta, \tau_p \gg \tau_d \\ \mathcal{O}(J\beta\tau_d\tau_p) & : J \gg \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\beta\tau_p^2) & : J \gg \beta, \tau_p \gg \tau_d \end{cases} \quad (94)$$

determined for RGA_{32c} , we can conclude that $RGA_{16a'}$ is not the desired sequence construction for interaction-dominated dynamics when $\tau_p > \tau_d$ due to the presence of $\mathcal{O}(J^2 \tau_p^2)$ terms in the effective error Hamiltonian [see Eq. (89)]. The concatenated construction designated by Eq. (93) alleviates this problem and achieves complete suppression of $\mathcal{O}(J^2 \tau_p^2)$ terms. In fact, RGA_{32c} turns out to be quite special in that additional sequence structures such as $RGA_{8a'}[RGA_4]$, $RGA_4[RGA_{8a'}]$, and $RGA_4[RGA_{8c}]$ are unable to attain such a high level of decoupling. Most notable are those which utilize RGA_4 at the outer layer of the concatenation and fall victim to the less desirable decoupling properties of the $K = 4$ sequences.

In the remaining regions of the parameter space both sequences scale equivalently, yet $RGA_{16a'}$ is found to be the optimal sequence. Comparing the effective error Hamiltonian

for $RGA_{16a'}$ given in Eq. (89) to that of RGA_{32c} given by

$$H_{\text{err}}^{RGA_{32c}} \approx \frac{16i\tau_p\tau_d}{\tau_c} \left(\sigma^x [B_0, B_x - \frac{2}{\pi} B_z] + \sigma^y [B_0, B_y - \frac{2}{\pi} B_x] + \sigma^z [B_0, B_z - \frac{2}{\pi} B_x + \frac{4}{\pi} B_y] \right), \quad (95)$$

the lowest order difference between the evolutions is essentially an additional term along the σ^z error channel for RGA_{32c} . Consequently, $RGA_{16a'}$ achieves a higher performance than RGA_{32c} when $\beta > J$ and in the strong-pulse regime when $\beta < J$. Note that $RGA_{16a'}$ is also optimal for a small region of the parameter space when $J \gg \beta$ and $\tau_p \gg \tau_d$. This sudden appearance of $RGA_{16a'}$ occurs as $J\tau_d \rightarrow 1$, where the $\mathcal{O}(J^2 \tau_p^2)$ scaling is more favorable than the $\mathcal{O}(J^3 \tau_p^3)$ scaling for RGA_{32c} .

The most interesting aspect of the $K = 32$ analysis is that RGA_{32b} achieves error reduction, yet the performance remains $\mathcal{O}(\tau_p \tau_d)$ in the strong pulse regime indicating a possible saturation in performance and significant error suppression capabilities. This artifact is due to the presence of pulse imperfections in the effective error Hamiltonian that may not be suppressed further simply by adjusting the sequence configuration. For this reason, it may be necessary to consider pulse shaping techniques to effectively average out additional finite-width contributions [72–74].

5. 64-Pulse Sequences

In the case of $K = 64$, robust DD is mainly characterized by the single layer concatenation of RGA_{8c} :

$$RGA_{64c} = RGA_{8c}[RGA_{8c}] \quad (96)$$

and $4 \times RGA_{16a'}$. The distinction between the sequences is evident from Fig. 3(e), where RGA_{64c} is preferred in the interaction-dominant regime primarily when $\tau_p \gg \tau_d$ and $4 \times RGA_{16a'}$ is optimally robust when the environment dynamics dominate. The partitioning of the space can be understood by comparing the performance scaling of $RGA_{16a'}$ given in Eq. (88) and

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_d\tau_p) & : J \ll \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\beta\tau_p^2) & : J \ll \beta, \tau_p \gg \tau_d \\ \mathcal{O}(J\beta\tau_d\tau_p) & : J \gg \beta, \tau_p \ll \tau_d \\ \mathcal{O}(J\beta\tau_p^2) & : J \gg \beta, \tau_p \gg \tau_d \end{cases} \quad (97)$$

determined for RGA_{64c} , most notably when $J \gg \beta$ and $\tau_d \gg \tau_p$, where terms of $\mathcal{O}(J^2 \tau_p^2)$ are effectively suppressed for RGA_{64c} . For $J \ll \beta$, it is not possible to draw any conclusions regarding sequence preference from the scaling equations since both exhibit identical performance scalings. Therefore, we turn to the effective Hamiltonian to gain some additional insight.

For the case of $RGA_{16a'}$, the effective error Hamiltonian is presented in Sec. IV B 3 [see Eq. (89)], while for RGA_{64c} we find

$$H_{\text{err}}^{RGA_{64c}} \approx \frac{32i\tau_d\tau_p}{\tau_c} \left(\sigma^x [B_0, B_x + \frac{2}{\pi} B_y] + \sigma^y [B_0, B_y - \frac{2}{\pi} B_z] + \sigma^z [B_0, B_z + \frac{2}{\pi} B_y] \right). \quad (98)$$

Interestingly, both sequences have effective error Hamiltonians that contain similar low-order terms in the $\beta \ll J$ regime. Substantial differences appear in higher order terms of $\mathcal{O}(\beta^2 J \tau_d^2 \tau_p)$ and $\mathcal{O}(\beta^2 J \tau_d \tau_p^2)$, where $RGA_{16a'}$ generates far less effective decoherence than RGA_{64c} . This result is consistent with the values of sequence performance which tend to deviate more strongly as β increases beyond J . In the interaction-dominant regime, the advantage of RGA_{64c} over additional sequences, e.g., 2 cycles of RGA_{32c} , is the increased suppression of $\mathcal{O}(J\beta\tau_p^2)$ and $\mathcal{O}(J\beta\tau_p\tau_d)$ evident in the effective error Hamiltonian when $\tau_p > \tau_d$ (not shown) and $\tau_p < \tau_d$ [compare Eq. (95) and Eq. (98)], respectively. A similar performance and effective Hamiltonian analysis also conveys the inadequacies of $RGA_{8a'}$ [$RGA_{8a'}$] for suppressing $\mathcal{O}(J\beta\tau_d\tau_p)$ and $\mathcal{O}(J^2\tau_p^2)$ terms.

6. 256-Pulse Sequences

The space of optimal sequences for $K = 256$ is characterized by previously identified sequences: $16 \times RGA_{16a'}$ and $4 \times RGA_{64c}$ with performance scalings given by Eqs (88) and (97) and effective error Hamiltonians presented in Eqs. (89) and (98), respectively. In Fig. 3(f), we present a comprehensive numerical summary for the regions of optimal performance for $K = 256$ as a function of J and τ_p for fixed β and τ_d . Note the similarities between the $K = 256$ and $K = 64$ numerical results which are nearly identical, as $RGA_{16a'}$ is optimal for $J < \beta$ and RGA_{64c} is favorable for $J > \beta$, due to the preference in optimal sequence structure. More intricate sequence constructions such as

$$RGA_{256c} := RGA_4[RGA_{64c}] \quad (99)$$

suffer from an accumulation of $\mathcal{O}(J\tau_p)$ terms generated by the concatenation with RGA_4 on the outer level. A performance scaling similar to that of RGA_{64c} can be re-established by reversing the concatenation order according to $RGA_{64c}[RGA_4]$, yet an accumulation of $\mathcal{O}(J\beta\tau_d\tau_p)$ and $\mathcal{O}(J^2\tau_p^2)$ terms renders this sequence structure less favorable than $4 \times RGA_{64c}$.

A summary of the performance scaling equations for all optimal sequences discussed in Sec. IV B is presented in Table III. Note that first order error suppression, in τ_p , is achieved for a majority of K_{opt} . However, we are only able to demonstrate the reduction of second order decoherence operators, such as the suppression of $\mathcal{O}(J^2\tau_d\tau_p)$ terms for certain cases, and not complete suppression of $\mathcal{O}(\tau_p\tau_d)$ or $\mathcal{O}(\tau_p^2)$ terms. This result is consistent with DD no-go theorems

Sequence	$J \ll \beta$		$J \gg \beta$	
	$\tau_p \ll \tau_d$	$\tau_p \gg \tau_d$	$\tau_p \ll \tau_d$	$\tau_p \gg \tau_d$
RGA_4	$\mathcal{O}(J\beta\tau_d^2)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J^2\tau_d^2)$	$\mathcal{O}(J\tau_p)$
$RGA_{4'}$	$\mathcal{O}(J\beta\tau_d^2)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J^2\tau_d^2)$	$\mathcal{O}(J\tau_p)$
$RGA_{8a'}$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_p^2)$	$\mathcal{O}(J^2\tau_d\tau_p)$	$\mathcal{O}(J^2\tau_p^2)$
RGA_{8b}	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$
RGA_{8c}	$\mathcal{O}(J\beta\tau_d^2)$	$\mathcal{O}(J\beta\tau_p^2)$	$\mathcal{O}(J^2\tau_d^2)$	$\mathcal{O}(J^2\tau_p^2)$
$RGA_{16a'}$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J^2\tau_p^2)$
$RGA_{16b'}$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J^2\tau_p^2)$
RGA_{32c}	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_p^2)$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_p^2)$
RGA_{64c}	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_p^2)$	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\beta\tau_p^2)$
RGA_{256c}	$\mathcal{O}(J\beta\tau_d\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$

TABLE III. Summary of distance measure (D) scalings for each optimal RGA_K sequence located by our search algorithm, for DD evolution subjected to finite-width rectangular pulses of duration τ_p pulses and pulse-interval τ_d .

which prove that it is not possible to suppress decoherence operators that are manifested by the second order perturbation expansion for the pulse error evolution operator, i.e. $\mathcal{O}(\tau_p\tau_d)$ and $\mathcal{O}(\tau_p^2)$ terms, when rectangular pulse profiles are utilized [75, 76]. Our analysis is consistent with these theorems and further conveys the need to utilize pulse shaping techniques in conjunction with optimal sequence construction to achieve high order error suppression in the presence of finite-width pulses. Indeed, when liberated from the constraint of rectangular pulse profiles, pulse sequences using DCG and CDCG [46–48] may be employed when pulse-width errors are the dominant concern.

C. Flip-angle errors

An additional form of pulse error we consider is that of a flip-angle error. In Fig. 5, we analyze the performance of DD sequences optimized for flip-angle errors with respect to the ratio $J/\beta \in [10^{-6}, 10^6]$ for $\beta = 1\text{kHz}$ and $\epsilon \in [0, 0.2]$. The results are averaged over 10 realizations of Eq. (15).

The control pulse set \mathcal{G} is equivalent to Eq. (70), with the pulse profile defined in Eq. (20). The resulting unitary pulse operators are given by

$$X(Y, Z) = e^{-i\pi/2(1+\epsilon)\sigma^x(y,z)} \quad (100)$$

and $\{\bar{X}, \bar{Y}, \bar{Z}\} = \{X^\dagger, Y^\dagger, Z^\dagger\}$. The analysis is symmetric with respect to over or under-rotations, therefore, our focus on over-rotations does not result in a loss of generality. As in the case of finite-width pulses, the initial population is considered to be only 16 of the 49 initial configurations chosen at random, each with an equal probability.

1. 4-Pulse Sequences

In the case of flip-angle errors, the optimal sequence construction for $K = 4$ is primarily $RGA_2 = P_2 f_{\tau_d} P_2 f_{\tau_d}$. The

performance can be shown to scale according to

$$D \sim \begin{cases} \mathcal{O}(J\tau_d) & : J \ll \beta, \epsilon \ll J\tau_d \\ \mathcal{O}(\epsilon J\tau_d) & : J \ll \beta, \epsilon \gg J\tau_d \\ \mathcal{O}(J\tau_d) & : J \gg \beta, \epsilon \ll J\tau_d \\ \mathcal{O}(\epsilon J\tau_d) & : J \gg \beta, \epsilon \gg J\tau_d \end{cases}, \quad (101)$$

which is in direct agreement with the effective error Hamiltonian:

$$\begin{aligned} \bar{H}_{\text{err}}^{RGA_2} &\approx \sigma^x B_x + \frac{\pi^2 \epsilon^2}{4} (\sigma^y B_y + \sigma^z B_z) \\ &\quad - \frac{\pi \epsilon}{2} (\sigma^y B_z - \sigma^z B_y). \end{aligned} \quad (102)$$

In Fig. 5(a), the performance for RGA_2 is shown as a function of ϵ and the ratio J/β . RGA_2 is the optimal sequence for all values of ϵ and J within the designated parameter range, except for a particular region at $J/\beta = 10^5$, where RGA_4 is preferred. Analyzing the effective error Hamiltonian for RGA_4 in the presence of flip-angle errors, we find

$$\begin{aligned} \bar{H}_{\text{err}}^{RGA_4} &\approx -\frac{\pi^2 \epsilon^2}{8\tau_d} \sigma^z - \frac{\pi \epsilon}{2} \sigma^z B_{x-y} \\ &\quad + \frac{\pi^2 \epsilon^2}{4} [\sigma^x B_{x-y} + \sigma^y (B_y - 2B_x)], \end{aligned} \quad (103)$$

which contains an additional error term of $\mathcal{O}(\epsilon^2/\tau_d)$. An examination of the performance yields

$$D \sim \begin{cases} \mathcal{O}(J\beta\tau_d^2) & : J \ll \beta, \epsilon \ll J\tau_d \\ \mathcal{O}(\epsilon^2) & : J \ll \beta, \epsilon \gg J\tau_d \\ \mathcal{O}(J^2\tau_d^2) & : J \gg \beta, \epsilon \ll J\tau_d \\ \mathcal{O}(\epsilon^2) & : J \gg \beta, \epsilon \gg J\tau_d \end{cases}, \quad (104)$$

where the term quadratic in ϵ shown in Eq. (103) indeed dictates the scaling of the performance in the $\epsilon \gg J\tau_d$ regime. Since the magnitude of the flip-angle error is determined by the accuracy of the experimental equipment and cannot be directly manipulated, terms which are not proportional to the delay time are significantly more destructive to DD evolution. An important question to be addressed is whether high order error suppression is attainable for errors which are solely proportional to the flip-angle error, or if the amplitude of such error terms can be reduced dramatically, simply by sequence construction.

2. 8-Pulse Sequences

We identify

$$RGA_{8a} = IP_1 P_2 \bar{P}_1 IP_1 \bar{P}_2 P_1 \quad (105)$$

as the only optimal sequence for $K = 8$ in the presence of flip-angle errors. The structure is similar to GA_{8a} and $RGA_{8a'}$, differing only by pulse phases, and identical to a time-symmetrized version of RGA_4 sequence, namely $RGA_4 RGA_4$. Time-symmetrization has long been known to

be beneficial for DD sequence construction since all odd-order terms in the effective error Hamiltonian are averaged out [25, 77], even in the case of pulse errors [13]. The effect of symmetrization is apparent from the significant increase in performance amplitude, implying a reduction in error terms $\mathcal{O}(\epsilon^2)$; see Fig. 5(b). In fact, a direct analysis of the scaling of the performance reveals complete cancellation of $\mathcal{O}(\epsilon^2)$ and

$$D \sim \mathcal{O}(\epsilon J\tau_d) \quad \forall J, \beta, \epsilon, \tau_d. \quad (106)$$

The scaling is supported by the effective error Hamiltonian for RGA_{8a} :

$$\begin{aligned} \bar{H}^{RGA_{8a}} &\approx -\frac{\pi \epsilon}{2} \sigma^z B_{x-y} + \frac{\pi^2 \epsilon^2}{4} \sigma^y B_{x-y} \\ &\quad - \frac{\pi^2 \epsilon^2}{4} \sigma^y (2B_x - B_y), \end{aligned} \quad (107)$$

where the $\mathcal{O}(\epsilon J\tau_d)$ term is accompanied by two additional $\mathcal{O}(\epsilon^2 J\tau_d)$ terms. In order to provide successive error suppression of terms $\mathcal{O}(\epsilon^n)$, $n \geq 1$, it is necessary to continue producing dominant decoherence operators that are linear, or even quadratic, in $J\tau_d$ to effectively control the amplitude of the errors associated with ϵ .

3. 16-Pulse Sequences

Optimal performance for $K = 16$ is obtained for a robust version of GA_{16a} denoted by

$$RGA_{16a} = \bar{P}_3 (RGA_{8a}) P_3 (RGA_{8a}). \quad (108)$$

The structure utilizes a robust version of the generalized CPMG sequence, RGA_2 , to encompass the RGA_{8a} sequence and achieve a performance scaling

$$D \sim \mathcal{O}(\epsilon^2 J\tau_d) \quad \forall J, \beta, \epsilon, \tau_d. \quad (109)$$

The additional level of concatenation enhances robustness and results in a quadratic improvement in error suppression over RGA_{8a} . An analysis of the performance for RGA_{16a} as a function of ϵ and the ratio J/β is provided in Fig. 5(c).

In constructing the RGA_{16a} sequence there is a freedom in the choice of P_3 . Similar to our discussion in Sec. IV B 2 regarding the possible sequence variations of RGA_{8b} , let us consider $\{P_1, P_2\} = \{X, Y\}$ and analyze the effective error Hamiltonian for each possible non-trivial P_3 variation. Taking $P_3 = X, Y, Z$, we obtain the following effective error Hamiltonians:

$$\bar{H}_{\text{err}, P_3=X}^{RGA_{16a}} \approx \frac{\pi^2 \epsilon^2}{4} [\sigma^x B_{x-y} - \sigma^y B_{x-y}], \quad (110a)$$

$$\bar{H}_{\text{err}, P_3=Y}^{RGA_{16a}} \approx \frac{\pi^2 \epsilon^2}{4} [\sigma^x B_{x-y} - \sigma^y (2B_x - B_y)], \quad (110b)$$

$$\bar{H}_{\text{err}, P_3=Z}^{RGA_{16a}} \approx \frac{\pi \epsilon}{2} \sigma^z B_{x-y}. \quad (110c)$$

The first two cases remain quadratic in ϵ and consistent with the scaling of Eq. (109). The last case exhibits a diminished error suppression and scales as $\mathcal{O}(\epsilon J\tau_d)$, the same as

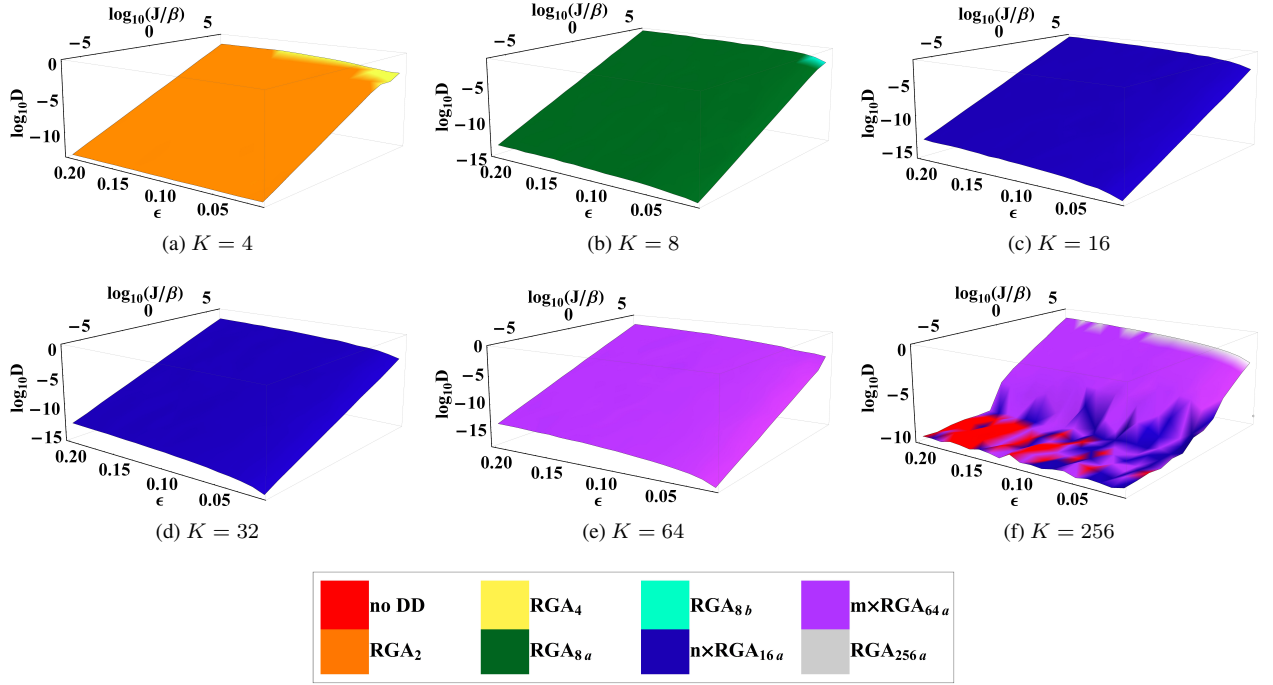


FIG. 5. (color online) Performance of RGA_K sequences for $K = 4, 8, 16, 32, 64, 256$, as shown in (a)-(f), respectively, as a function of J/β and ϵ . The minimum pulse interval is fixed at $\tau_d = 0.1\text{ns}$, $J/\beta \in [10^{-6}, 10^6]$, and $\beta = 1\text{kHz}$, while ϵ is varied from one to twenty percent rotation error. Results are averaged over 10 realizations of B_μ . Sequence performance mostly increases from $K = 4$ to $K = 64$, indicating a reduction in the error terms proportional to ϵ in the effective error Hamiltonian. Successive error suppression is achieved for $K = 4, 8, 16, 64$, where the maximum error suppression yields $D \sim \mathcal{O}(\epsilon^3 J \tau_d)$ for RGA_{64a} . Multiple cycles of RGA_{16a} and RGA_{64a} appear as optimal sequences for various K_{opt} . The number of cycles for each sequence is given as follows: (d) $n = 2$, (e) $m = 1$, and (f) $n = 16$, $m = 4$.

RGA_{8a} . Of course, this can be understood from Eq. (107) where the term linear in ϵ resides along the σ^z channel. Applying a Z pulse commutes with this term and, therefore, does not provide any additional error suppression. As previously discussed, choosing the correct sequence configuration is critical in constructing a robust sequence. This is especially true in cases like RGA_{16a} where the outer-level concatenating sequence does not address all error channels, i.e., the pulse operators do not anti-commute with all possible system-environment interaction terms.

4. 32-Pulse Sequences

The sequence structures obtained so far for robustness against flip-angle errors suggests that a robust version of GA_{32a} would appear as the optimal sequence for $K = 32$. In actuality, robust versions of GA_{32a} do not exhibit ample increases in performance over RGA_{16a} . The sequence construction

$$RGA_{32a} = RGA_4[RGA_{8a}] \quad (111)$$

can be shown to yield a performance scaling

$$D \sim \begin{cases} \mathcal{O}(\epsilon^2) & : J \ll \beta, \forall \epsilon, \tau_d \\ \mathcal{O}(\epsilon J \beta \tau_d^2) & : J \gg \beta, \epsilon \ll J \tau_d \\ \mathcal{O}(\epsilon^2) & : J \gg \beta, \epsilon \gg J \tau_d \end{cases} \quad (112)$$

and generate an effective error Hamiltonian

$$\bar{H}_{\text{err}}^{RGA_{32a}} \approx -\frac{\pi^2 \epsilon^2}{64 \tau_d} \sigma^z + 2\pi i \epsilon \tau_d \sigma^z [B_0, B_{x-y}], \quad (113)$$

both of which display a dominant decoherence-generating term of $\mathcal{O}(\epsilon^2)$. The preferred sequence structure for $K = 32$ is RGA_{16a} , where two cycles are applied to generate a 32-pulse sequence. Saturation in performance is perhaps an indication that maximum robustness against flip-angle errors has been obtained and attempting to apply more sophisticated sequence structures simply results in error accumulation, rather than suppression. Error accumulation is validated here with the reemergence of the $\mathcal{O}(\epsilon^2)$ term in Eq. (113), that had been previously suppressed by RGA_{8a} . Furthermore, the reduced performance can be attributed to the outer layer sequence, RGA_4 , which generates $\mathcal{O}(\epsilon^2)$ terms that essentially destroy the additional error suppression achieved by time-symmetrization. One way of overcoming additional decoherence brought about by flip-angle errors is to use composite pulses which essentially mimic the application of a single π -pulse by applying a number of variable angled pulses in succession [33, 54]. A summary of $K = 32$ performance is given in Fig. 5(d), where $n \times RGA_{16a}$ with $n = 2$ denotes two successive cycles of RGA_{16a} .

5. 64-Pulse Sequences

Robust DD is achieved at $K = 64$ by

$$RGA_{64a} = RGA_{8a}[RGA_{8a}], \quad (114)$$

a robust version of GA_{64a} . In contrast to $K = 32$, applying multiple (in this case four) cycles of RGA_{16a} does not attain optimal performance for $K = 64$. Examining the scaling of the performance, we find that RGA_{64a} attains an additional order of error suppression over RGA_{16a} that is immediately evident from

$$D \sim \mathcal{O}(\epsilon^3 J \tau_d) \quad \forall J, \beta, \epsilon, \tau_d \quad (115)$$

and

$$\bar{H}^{RGA_{64a}} \approx \frac{\pi^3 \epsilon^3}{4} \sigma^z B_y - \frac{3\pi^3 \epsilon^3}{8} \sigma^z B_x, \quad (116)$$

where the $\mathcal{O}(\epsilon^2 J \tau_d)$ term is effectively suppressed. In the previous discussion on $K = 32$ sequences, we noted that it is possible that the outer RGA_4 sequence diminished performance due to its lower order error suppression. The fact that an additional order of error suppression over $K = 16$ is achieved here for a concatenated structure comprising two RGA_{8a} sequences, both of which supply the same order of error suppression, further validates this claim. Therefore, one of the important aspects of sequence construction for concatenated protocols is to utilize sequences which at least obtain the same order of error suppression. Figure 5(e) summarizes the $K = 64$ performance over a range of ϵ and J/β values.

6. 256-Pulse Sequences

In the case of $K = 256$, the optimal sequence configurations appear as multi-cycle versions of previously identified sequences. Optimal performance in the system-environment interaction-dominant ($J \geq \beta$) regime is achieved by $4 \times RGA_{64a}$, while for $J < \beta$ the parameter space is shared by $16 \times RGA_{16a}$ and $4 \times RGA_{64a}$. As shown in Fig. 5(f), sequence performance eventually degrades further and it becomes more favorable to not apply DD at all for large ϵ .

Similar to $K = 32$, robust versions of $K = 256$ sequences are not found to be optimal. Previously located sequences suggest that

$$RGA_{256a} = RGA_4[RGA_{64a}], \quad (117)$$

may serve as a robust configuration for flip-angle errors; however, this turns out to be false. We attribute the superiority of multi-cycle sequences to the outer-most level of concatenation of RGA_{256a} since it does not achieve the same decoupling order as the inner levels. Compared to RGA_{64a} , RGA_{256a} receives a substantial reduction in performance by the addition of the outer RGA_4 sequence and re-emerges with a distance measure scaling

$$D \sim \mathcal{O}(\epsilon^2) \quad \forall J, \beta, \epsilon, \tau_d, \quad (118)$$

Sequence	$J \ll \beta$		$J \gg \beta$	
	$\epsilon \ll J \tau_d$	$\epsilon \gg J \tau_d$	$\epsilon \ll J \tau_d$	$\epsilon \gg J \tau_d$
RGA_2	$\mathcal{O}(J \tau_d)$	$\mathcal{O}(\epsilon J \tau_d)$	$\mathcal{O}(J \tau_d)$	$\mathcal{O}(\epsilon J \tau_d)$
RGA_4	$\mathcal{O}(J \beta \tau_d^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(J^2 \tau_d^2)$	$\mathcal{O}(\epsilon^2)$
RGA_{8a}	$\mathcal{O}(\epsilon J \tau_d)$	$\mathcal{O}(\epsilon J \tau_d)$	$\mathcal{O}(\epsilon J \tau_d)$	$\mathcal{O}(\epsilon J \tau_d)$
RGA_{16a}	$\mathcal{O}(\epsilon^2 J \tau_d)$	$\mathcal{O}(\epsilon^2 J \tau_d)$	$\mathcal{O}(\epsilon^2 J \tau_d)$	$\mathcal{O}(\epsilon^2 J \tau_d)$
RGA_{32a}	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$
RGA_{64a}	$\mathcal{O}(\epsilon^3 J \tau_d)$	$\mathcal{O}(\epsilon^3 J \tau_d)$	$\mathcal{O}(\epsilon^3 J \tau_d)$	$\mathcal{O}(\epsilon^3 J \tau_d)$
RGA_{256a}	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$

TABLE IV. Summary of distance measure D scalings for each optimal RGA_K sequence located for DD pulses subjected to flip-angle errors with rotation error ϵ with fixed pulse-interval τ_d .

which is consistent with that of RGA_4 given in Eq. (104) for certain parameter regimes. The effective error Hamiltonian

$$\begin{aligned} \bar{H}^{RGA_{256a}} \approx & -\frac{\pi^2 \epsilon^2}{512 \tau_d} \sigma^z + \frac{\pi^3 \epsilon^3}{1024 \tau_d} (\sigma^x - \sigma^y) \\ & + 4\pi^3 i \epsilon^3 \tau_d \sigma^z [B_0, 3B_x - 2B_y] \end{aligned} \quad (119)$$

further substantiates the numerically-determined performance scaling. In Table IV, we display the performance scaling for $K = 256$ along with the scalings for the remaining values of K_{opt} discussed above.

In summary, the results presented for $K=K_{\text{opt}}$ in the presence of flip-angle errors demonstrate that successive error suppression is achievable by concatenation only if the outer sequence maintains the same decoupling order as the inner sequence(s). Supplying a lower order decoupling outer sequence ultimately leads to an effective error Hamiltonian that possesses dominant error terms that are intrinsic to the low-order sequence. Such a situation did not arise for finite-width pulse errors due to the absence of pulse shaping. Uninhibited by no-go theorems for flip-angle errors, we believe that extending the search beyond $K = 256$ will result in additional sequence configurations that utilize concatenations of RGA_{8a} or even a more robust construction such as RGA_{16a} . This conclusion is supported by the fact that the performance scaling obtained for $RGA_{512a} = RGA_{8a}[RGA_{8a}[RGA_{8a}]]$ is $D \sim \mathcal{O}(\epsilon^5 J \tau_d)$, which suggests that arbitrary order error suppression using ℓ concatenations of RGA_{8a} can be used to achieve $D \sim \mathcal{O}(\epsilon^{2\ell-1} J \tau_d)$.

D. Finite-Width and Flip-Angle Errors

Flip-angle and finite-width pulse errors are prevalent in a variety of experimental settings. Therefore, it is necessary for DD sequences to be robust against both types of errors simultaneously if reliable computation is to be implemented under the protection of DD in realistic setups. Designs based on (C)DCG [46–48] do not apply in this case as they do not address flip-angle errors. The KDD sequence is applicable, but unlike the present study, it employs non- π pulses [54]. As a final consideration, we investigate the inclusion of the two forms of pulse errors and search for optimal sequences at each

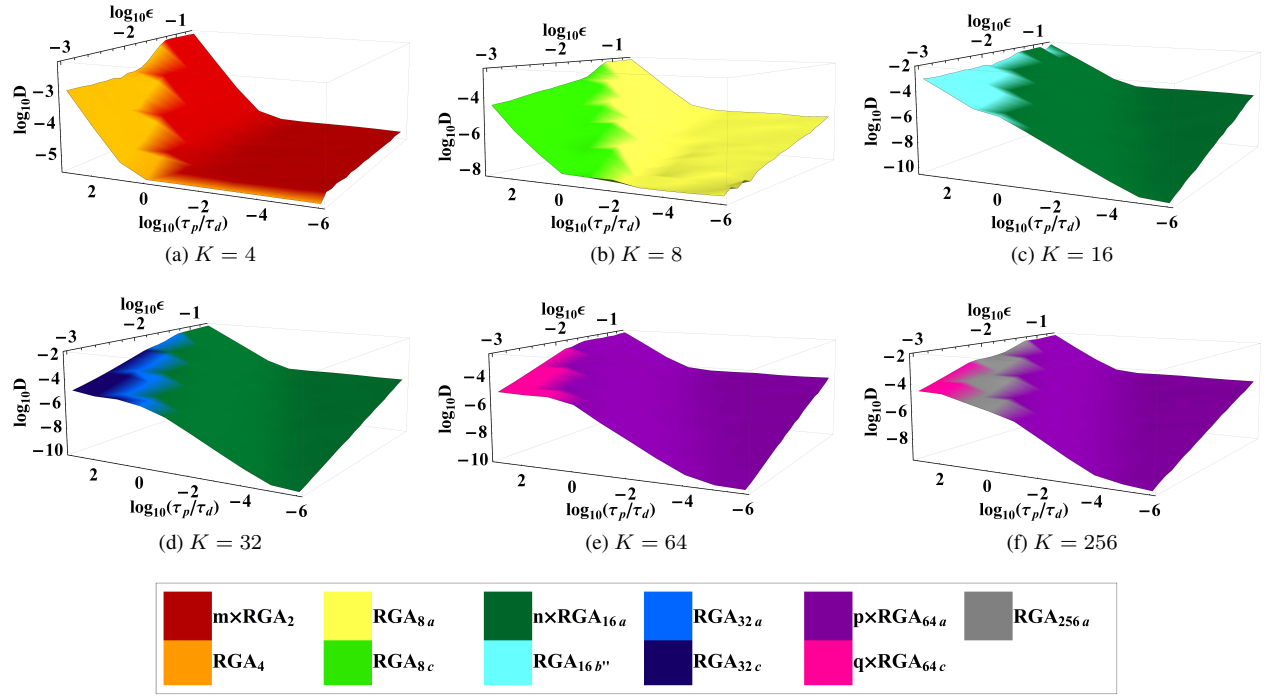


FIG. 6. (color online) Performance of RGA_K sequences for $K = 4, 8, 16, 32, 64, 256$, as shown in (a)-(f), respectively, as a function of ϵ and τ_p/τ_d . The minimum pulse-interval is fixed at $\tau_d = 0.1\text{ns}$, while $J = 1\text{MHz}$ and $\beta = 1\text{kHz}$. The flip-angle error is varied from a 1% to a 10% error and τ_p is varied throughout a wide range of values so that $\tau_p \ll \tau_d$, and $\tau_p \gg \tau_d$, is explored. The majority of the parameter space is dominated by the RGA_{K_a} sequences, even in the finite-width error-dominant regime. The robustness of RGA_{K_a} to pulse imperfections ultimately saturates at $K = 64$. All simulations are averaged over 10 realizations of B_μ . The optimal sequences that are used in multiple K_{opt} values require the following number of cycles: (a) $m = 2$, (c) $n = 1$, (d) $n = 2$, (e) $p = q = 1$, (f) $p = q = 4$.

K_{opt} . By performing this search, we are essentially addressing the possibility of constructing fault-tolerant DD in the most convenient possible arrangement: fixed pulse-interval and rectangular pulse shape.

The control pulse set \mathcal{G} again contains 7 possible pulses, where each pulse operator is defined by

$$X(Y, Z) = e^{-i\tau_p[A(1+\epsilon)\sigma^x(y, z) + H]} \quad (120)$$

and $\{\bar{X}, \bar{Y}, \bar{Z}\}$ correspond to 180-degree phase rotations, i.e. $A \rightarrow -A$. The amplitudes of the error and bath Hamiltonians are chosen as $J = 1\text{MHz}$ and $\beta = 1\text{kHz}$, respectively. Optimal sequence performance is analyzed with respect to ϵ and τ_p in the regime where $J \gg \beta$ and $J\tau_d \ll 1$ to characterize sequence robustness as a function of errors generated by the DD pulses.

1. 4-Pulse Sequences

The optimal sequence configurations for $K = 4$ are RGA_2 and RGA_4 . Since RGA_2 consists of only two pulses, the cycle is simply repeated twice to generate a four pulse sequence, i.e. $2 \times RGA_2$. In Fig. 6(a), $2 \times RGA_2$ is shown to be optimal for a considerable portion of the parameter range considered: $\tau_p < \tau_d$ when $\epsilon > J\tau_p$. Dominance is exchanged at $\tau_p = \tau_d$ and RGA_4 becomes the preferred sequence con-

figuration. In order to understand the variations in optimal sequence selection, we numerically determine the scaling of distance measure D and further characterize sequence performance by computing the effective error Hamiltonians for both sequences in the regime where $J\tau_d \ll 1$.

In the case of RGA_2 , the performance can be shown to scale according to

$$D \sim \begin{cases} \mathcal{O}(J\tau_c) & : \tau_p \ll \tau_d \forall \epsilon \\ \mathcal{O}(J\tau_p) & : \tau_p \gg \tau_d \forall \epsilon \end{cases}, \quad (121)$$

which is dictated by first order terms in either τ_d or τ_p . This result is quite different from RGA_4 , where

$$D \sim \begin{cases} \mathcal{O}(J^2\tau_d^2) & : \epsilon \ll J\tau_p \ll J\tau_d \\ \mathcal{O}(J^2\tau_d^2) & : J\tau_p \ll \epsilon \ll J\tau_d \\ \mathcal{O}(\epsilon^2) & : J\tau_p \ll J\tau_d \ll \epsilon \\ \mathcal{O}(J\tau_p) & : \epsilon \ll J\tau_d \ll J\tau_p \\ \mathcal{O}(J\tau_p) & : J\tau_d \ll \epsilon \ll J\tau_p \\ \mathcal{O}(\epsilon^2) & : J\tau_d \ll J\tau_p \ll \epsilon \end{cases} \quad (122)$$

indicates a more complicated scaling structure. Considering the case where $J\tau_p \ll J\tau_d \ll \epsilon$, which is essentially captured by $\epsilon = 10^{-3}$ in the strong pulse regime [$\tau_p \ll \tau_d$] in Fig. 6(a), RGA_2 is the preferred sequence configuration due to its scaling in τ_p rather than ϵ^2 . Performing a similar analysis for

$J\tau_d \ll J\tau_p$ with $\epsilon \ll J\tau_p$ is more complicated since both sequences scale equivalently: $D \sim \mathcal{O}(J\tau_p)$. Consequently, we employ average Hamiltonian theory to elucidate the details.

The evolution generated by RGA_2 results in the effective error Hamiltonian

$$\bar{H}_{\text{err}}^{RGA_2} \approx \sigma^x B_x - \frac{4\tau_p}{\pi\tau_c}(1-\epsilon)\sigma^z B_y - \frac{2\epsilon\tau_p}{\tau_c}\sigma^y(B_y - \frac{2}{\pi}B_z) - \frac{\pi\epsilon\tau_d}{\tau_c}(\sigma^y B_z - \sigma^z B_y), \quad (123)$$

while for RGA_4

$$\bar{H}_{\text{err}}^{RGA_4} \approx -\frac{\pi^2\epsilon^2}{2\tau_c}\sigma^z - \frac{4\tau_p}{\pi\tau_c}\sigma^z B_{x-y} - \frac{2\epsilon\tau_p}{\tau_c}(\sigma^x - \sigma^y)B_{x-y} - \frac{4\epsilon\tau_p}{\pi\tau_c}\sigma^z \left[\left(1 + \frac{\pi^2}{4} + \frac{\pi^2\tau_d}{2\tau_p}\right)B_{x-y} + \frac{\pi}{2}B_y \right] \quad (124)$$

The application of two cycles of RGA_2 to match the four-pulse sequence of RGA_4 leads to a factor of 2 increase for the RGA_2 effective error Hamiltonian over that of RGA_4 , however, the implication is not necessarily that the performance will degrade by a factor of two. An additional possibility for RGA_4 -dominance in the $\epsilon \ll J\tau_p$ regime is the presence of the difference operator B_{x-y} in $\bar{H}_{\text{err}}^{RGA_4}$ rather than the single bath operator proportional to τ_p expressed in Eq. (123). RGA_4 performance appeared to benefit from the difference operator in the case of finite-width pulses so it is possible that the same situation exists here. The effective error Hamiltonians can be utilized to further provide consistency between the remaining numerically-calculated distance scalings and analytical expression, e.g., $\bar{H}_{\text{err}}^{RGA_2}$ conveys a suppression of $\mathcal{O}(\epsilon^2)$ terms, while $\bar{H}_{\text{err}}^{RGA_4}$ does not.

2. 8-Pulse Sequences

Two of the three robust sequence constructions identified in previous sections, RGA_{8a} and RGA_{8c} , characterize the parameter space for combined pulse errors. The most favorable sequence is found to be RGA_{8a} , as shown in Fig. 6(b), where it is optimal for all values of τ_p and ϵ except $\tau_p \gg \tau_d$ when $\epsilon \ll 1$. The remaining region is dominated by RGA_{8c} as expected by the finite-width error analysis presented above. Focusing on the interplay between errors generated by the DD pulses, the performance for RGA_{8a} is found to scale as

$$D \sim \begin{cases} \mathcal{O}(\epsilon J\tau_d) & : J\tau_p \ll \epsilon \ll J\tau_d \\ \mathcal{O}(\epsilon J\tau_d) & : J\tau_p \ll J\tau_d \ll \epsilon \\ \mathcal{O}(J\tau_p) & : \text{otherwise} \end{cases}, \quad (125)$$

while for RGA_{8c}

$$D \sim \begin{cases} \mathcal{O}(J^2\tau_d^2) & : \epsilon \ll J\tau_p \ll J\tau_d \\ \mathcal{O}(J^2\tau_d^2) & : J\tau_p \ll \epsilon \ll J\tau_d \\ \mathcal{O}(\epsilon^3) & : J\tau_p \ll J\tau_d \ll \epsilon \\ \mathcal{O}(J^2\tau_d\tau_p) & : \epsilon \ll J\tau_d \ll J\tau_p \\ \mathcal{O}(\epsilon J\tau_p) & : J\tau_d \ll \epsilon \ll J\tau_p \\ \mathcal{O}(\epsilon J\tau_p) & : J\tau_d \ll J\tau_p \ll \epsilon \end{cases}. \quad (126)$$

Using the scaling equations, we can attribute the dominance of RGA_{8a} when $J\tau_p \ll J\tau_d \ll \epsilon$ to the preference of $\mathcal{O}(\epsilon J\tau_d)$ scaling, as opposed to cubic scaling in ϵ found for RGA_{8c} , in the range of ϵ considered in Fig. 6(b). Taking $\epsilon < 10^{-3}$ [or in general $\epsilon^2 < J\tau_d$], the cubic scaling becomes more favorable, as one may expect, until $\epsilon \sim J\tau_d$ where the quadratic scaling in $J\tau_d$ begins to contribute significantly to the RGA_{8c} scaling. Equations (125) and (126) offer additional justification for the selection of RGA_{8c} as the optimal sequence when $J\tau_d \ll \epsilon \ll J\tau_p$ since RGA_{8c} has the ability to achieve $\mathcal{O}(J\tau_p)$ error suppression. The distance measure scalings for RGA_{8a} and RGA_{8c} can be further utilized to substantiate the preferred optimal sequence construction in the remaining regions displayed in Fig. 6(b).

The effective error Hamiltonians for the optimal $K = 8$ sequences, in the presence of finite-width and flip-angle errors, are essentially characterized by

$$\bar{H}_{\text{err}}^{RGA_{8a}} \approx \frac{4\tau_p}{\pi\tau_c}[\sigma^x B_z + \sigma^z(B_x - 2B_z)] - \frac{4\pi\epsilon\tau_d}{\tau_c}\sigma^z B_{x-y} \quad (127)$$

for RGA_{8a} and

$$\bar{H}_{\text{err}}^{RGA_{8c}} \approx (\sigma^x + \sigma^y) \left[\frac{\pi^3\epsilon^3}{2\tau_c} - \frac{4\epsilon\tau_p}{\tau_c}(B_x + B_y) \right] \quad (128)$$

for RGA_{8c} . The dynamics generated by RGA_{8a} focus primarily on the errors generated solely by the flip-angle error and do not attain suppression of decoherence operators proportional to both ϵ and τ_p . In contrast, RGA_{8c} focuses on both errors simultaneously, achieving first order suppression in τ_p and second order decoupling for errors solely proportional to ϵ . RGA_{8c} exhibits many attractive attributes for joint robustness against flip-angle and finite-width pulse errors, however, these features are constrained to sufficiently small ϵ , namely $\epsilon^2 < J\tau_d$. For the parameters considered here, this corresponds to a flip-angle error of $\epsilon < 0.1\%$ which may be quite ambitious for current experimental systems that can exhibit a rotation error well over 0.1% [17, 30, 66].

3. 16-Pulse Sequences

Robust decoupling is achieved for $K = 16$ using RGA_{16a} , where performance can be shown to scale as

$$D \sim \mathcal{O}(J\tau_p) \quad \forall \epsilon, \tau_p, \tau_d. \quad (129)$$

In the flip-angle error dominant regime, RGA_{16a} performance does not appear to be linearly or quadratically dependent on ϵ , implying the suppression of $\mathcal{O}(\epsilon^n)$ terms, for $n = 1, 2$. RGA_{16a} does not acquire robust decoupling against finite-width pulses and maintains the first order contribution in τ_p . As shown in Fig. 6(c),

$$RGA_{16b''} = RGA_{4'}[RGA_{4'}] \quad (130)$$

also appears as an optimal sequence in the regime where $\tau_p \gg \tau_d$ and $\epsilon < J\tau_p$. This sequence construction differs

slightly from $RGA_{16b'}$ in that both levels of the concatenation are composed of $RGA_{4'}$; hence, the double prime notation. The additional configurations, namely, RGA_{16b} and $RGA_{16b''}$, are not optimal in this case due to their inability to suppress decoherence operators linear in ϵ . In contrast, the performance for $RGA_{16b''}$ scales according to

$$D \sim \begin{cases} \mathcal{O}(\epsilon^2) & : \epsilon \gg J\tau_p \text{ and } \epsilon \gg J\tau_d \\ \mathcal{O}(J\tau_p) & : \text{otherwise} \end{cases}, \quad (131)$$

where the superiority of RGA_{16a} over $RGA_{16b'}$ is understood by the quadratic scaling in ϵ in the flip-angle dominant regime. In the regime where finite-width errors are prevalent, optimal robustness is obtained by $RGA_{16b''}$, even though both sequences exhibit the same scaling. We compute the effective error Hamiltonians for each sequence to determine the origin of $RGA_{16b''}$ robustness.

Let us first focus on the RGA_{16a} and consider the three possible sequence variations:

$$\bar{H}_{\text{err}, P_3=X}^{RGA_{16a}} \approx \frac{4\tau_p}{\pi\tau_c} (2\sigma^x B_z - \sigma^z B_y) \quad (132a)$$

$$\bar{H}_{\text{err}, P_3=Y}^{RGA_{16a}} \approx \frac{4\tau_p}{\pi\tau_c} (\sigma^z B_x - \sigma^x B_z) \quad (132b)$$

$$\bar{H}_{\text{err}, P_3=Z}^{RGA_{16a}} \approx \frac{4\tau_p}{\pi\tau_c} [\sigma^x B_y - \sigma^y B_x + 2\sigma^z (B_x - 2B_y)] \quad (132c)$$

for $\{P_1, P_2\} = \{X, Y\}$. Suppression of terms quadratic in ϵ is accomplished by all variations, consistent with Eq. (129). The distinction between the sequences lies in their ability to suppress $\mathcal{O}(J\tau_p)$ terms, with $P_3 = Y$ attaining the most suppression. In comparison, the suppression of decoherence operators generated by the finite duration of the pulses is more considerable for

$$\bar{H}_{\text{err}}^{RGA_{16b''}} \approx -\frac{\pi^2\epsilon^2}{2\tau_c}\sigma^y + \frac{4\tau_p}{\pi\tau_c}\sigma^y B_x \quad (133)$$

since only a single term of $\mathcal{O}(J\tau_p)$ exists. Therefore, $RGA_{16b''}$ is the preferred sequence due to its higher degree of error reduction, rather than suppression.

4. 32-Pulse Sequences

Degradation in robustness is first observed for $K = 32$, where 2 cycles of RGA_{16a} are found to be a more favorable sequence configuration than any other previously identified 32-pulse sequence for most of the parameter range considered. This result can be explained in terms of the dominant error type, whether it is flip-angle or finite pulse duration. In the $\epsilon \gg J\tau_p$ regime, RGA_{16a} is the optimal sequence due to its robustness to flip-angle errors. As the finite duration of the pulse becomes more detrimental, $\epsilon \ll J\tau_p$, RGA_{32a} performance surpasses that of RGA_{16a} . This is not surprising in light of the results shown in Sec. IV B 4, where we determined that RGA_{32a} provides first order suppression of error terms proportional to τ_p , while RGA_{16a} does not. The results

for $K = 32$ are summarized for a wide range of τ_p and ϵ in Fig. 6(d).

From a quantitative standpoint, $2 \times RGA_{16a}$ superiority can be deduced from the scaling of the distance measure D . The performance scaling for RGA_{16a} is given in Eq. (129), while for RGA_{32a} we determine that

$$D \sim \begin{cases} \mathcal{O}(\epsilon^2) & : J\tau_p \ll \epsilon, \forall \tau_d \\ \mathcal{O}(J\tau_p) & : J\tau_p \gg \epsilon, \forall \tau_d \end{cases}. \quad (134)$$

In the flip-angle error-dominant regime, the performance for RGA_{16a} remains proportional to τ_p while RGA_{32a} performance transitions from an $\mathcal{O}(J\tau_p)$ to an $\mathcal{O}(\epsilon^2)$ scaling. Consequently, the transition leads to RGA_{16a} becoming the more desirable sequence construction. Optimal sequence selection becomes harder to decipher in the $J\tau_p \gg \epsilon$ regime since both sequence scale equivalently. A comparison of the effective error Hamiltonian for RGA_{32a} ,

$$\bar{H}_{\text{err}}^{RGA_{32a}} \approx -\frac{\pi^2\epsilon^2}{2\tau_c}\sigma^z + \frac{4\tau_p}{\pi\tau_c}\sigma^z B_{x-y}, \quad (135)$$

and that of RGA_{16a} suggests that RGA_{32a} is more favorable when finite-width errors are dominant due to the difference operator B_{x-y} that is only distributed along a single error channel. This result is similar to the $K = 4$ case when only finite-width pulses are considered. RGA_4 emerges as the optimal sequence over $RGA_{4'}$ as a result of single-channel decoherence generated by B_{x-y} .

In addition to RGA_{32a} , RGA_{32c} also arises as an optimal sequence when $\tau_p \gg \tau_d$. RGA_{32c} 's region of optimal performance is highly dependent upon ϵ and mainly characterized by $\epsilon \ll 1$. From the performance scaling

$$D \sim \begin{cases} \mathcal{O}(\epsilon^2) & : \epsilon \gg J\tau_p, \forall \tau_d \\ \mathcal{O}(\epsilon J\tau_p) & : \epsilon \ll J\tau_p, \forall \tau_d \end{cases} \quad (136)$$

and effective error Hamiltonian

$$\bar{H}_{\text{err}}^{RGA_{32c}} \approx \frac{\pi^2\epsilon^2}{\tau_c}\sigma^x + \frac{4\epsilon\tau_p}{\tau_c}\sigma^y (B_y + B_z), \quad (137)$$

it is evident that RGA_{32c} attains first-order decoupling in τ_p and exhibits a performance scaling linear in ϵ when pulse-width errors are dominant. Hence, the flip-angle error must be relatively small ($\epsilon \ll 1$) in order for the $\mathcal{O}(\epsilon J\tau_p)$ scaling of RGA_{32c} to be advantageous.

5. 64-Pulse Sequences

Robustness to combined finite-width and flip-angle errors is regained at $K = 64$, where RGA_{64a} and RGA_{64c} are designated as the optimal sequence configurations. Similar to $K = 32$, we find that their regions of optimal performance are determined by the form of the dominant error. In Sec. IV C, RGA_{64a} was shown to be highly robust to flip-angle errors. Therefore, the $\epsilon > J\tau_p$ region utilizes RGA_{64a} as the optimal

sequence structure. The opposing region, where the finite duration of the pulse is the primary form of pulse error, utilizes RGA_{64c} due to its robustness against finite-width pulses.

Examining the performance as a function of τ_p and ϵ , we find that RGA_{64a} exhibits the scaling:

$$D \sim \mathcal{O}(J\tau_p) \quad \forall \epsilon, \tau_d, \tau_p \quad (138)$$

and the performance for RGA_{64c} is dictated by

$$D \sim \begin{cases} \mathcal{O}(\epsilon^2) & : J\tau_p \ll \epsilon, \forall \tau_d \\ \mathcal{O}(\epsilon J\tau_p) & : J\tau_p \gg \epsilon, \forall \tau_d \end{cases} \quad (139)$$

In the $\epsilon > J\tau_p$ regime, RGA_{64a} performance essentially scales unbiased by the finite-width contributions, as expected when flip-angle errors are dominant. RGA_{64c} becomes the more desirable sequence configuration when $\epsilon < J\tau_p$ due to its first order suppression in τ_p . The performance scaling obtained for both sequences are consistent with the results shown for $K = 64$ in Fig. 6(e).

In addition, Fig. 6(e) can also be explained in terms of the effective error Hamiltonians for RGA_{64a} and RGA_{64c} given by

$$\bar{H}_{\text{err}}^{RGA_{64a}} \approx \frac{4\tau_p}{\pi\tau_c} \sigma^x B_z + \frac{4\tau_p}{\pi\tau_c} \sigma^z (B_x - 2B_y) + \mathcal{O}(\epsilon\tau_p) \quad (140)$$

and

$$\bar{H}_{\text{err}}^{RGA_{64c}} \approx -\frac{\pi^2 \epsilon^2}{\tau_c} \sigma^y + \frac{4\epsilon\tau_p}{\tau_c} (\sigma^x B_x + \sigma^y B_z) \quad (141)$$

respectively. As expected from the distance measure scalings, $\bar{H}_{\text{err}}^{RGA_{64a}}$ is dominated by decoherence operators proportional to τ_p and $\bar{H}_{\text{err}}^{RGA_{64c}}$ is inhibited by a $\mathcal{O}(\epsilon^2)$ decoherence operator. One of the most interesting characteristics of RGA_{64a} , and all “a”-type sequences, is the persistent coefficient for the $\mathcal{O}(J\tau_p)$ term, despite the increase in cycle, an indication of error suppression for finite-width pulse errors. Another interesting artifact is the robustness of RGA_{64c} to $\mathcal{O}(\epsilon J\tau_p)$ errors. RGA_{64c} does not provide complete error suppression for these terms, yet there are significantly fewer terms present for RGA_{64c} than RGA_{64a} (not shown here).

6. 256-Pulse Sequences

Optimal performance is obtained primarily by four cycles of RGA_{64a} for $K = 256$. RGA_{256a} emerges as an optimal sequence construction when $J\tau_p > \epsilon$, however, is surpassed by the finite-width-optimized sequence $4 \times RGA_{64c}$ when $\tau_p \gg \tau_d$ and RGA_{64a} for the remainder of the parameter space shown in Fig. 6(f). Numerically analyzing the distance measure D as a function of ϵ , τ_p , and τ_d , the performance for RGA_{256a} can be shown to be characterized by the distance scaling

$$D \sim \begin{cases} \mathcal{O}(\epsilon^2) & : J\tau_p \ll \epsilon \\ \mathcal{O}(J\tau_p) & : J\tau_p \gg \epsilon \end{cases} \quad (142)$$

Comparing the performance scaling of RGA_{256a} to RGA_{64a} in the $\epsilon \gg J\tau_p$ regime, we find that RGA_{64a} is the more favorable sequence configuration due to its suppression of $\mathcal{O}(\epsilon^2)$ terms present in the effective error Hamiltonian. RGA_{256a} does not supply an equivalent suppression and its performance remains dictated by the quadratic term in this particular parameter regime. In the remaining regions of Fig. 6(e) the reason for RGA_{64a} selection over RGA_{256a} is indeterminable from the distance measure scalings since both are dictated by $\mathcal{O}(J\tau_p)$. We turn to the effective error Hamiltonian to provide the additional details.

The effective error Hamiltonian for RGA_{64a} is given by Eq. (140), while for RGA_{256a} we obtain

$$\begin{aligned} \bar{H}_{\text{err}}^{RGA_{256a}} \approx & -\frac{\pi^2 \epsilon^2}{2\tau_c} \sigma^z - \left[\frac{4\tau_p}{\pi\tau_c} \sigma^z + \frac{2\epsilon\tau_p}{\tau_c} (\sigma^x + \sigma^y) \right] B_{x-y} \\ & - \frac{4\epsilon\tau_p}{\pi\tau_c} \sigma^z \left[\left(1 - \frac{\pi^2}{4}\right) B_{x-y} - \frac{5\pi}{2} B_z \right]. \end{aligned} \quad (143)$$

Notice that both effective error Hamiltonians contain decoherence operators that are $\mathcal{O}(J\tau_p)$, however, RGA_{256a} contains only a single term. The additional terms in the effective error Hamiltonian for RGA_{64a} clearly result in RGA_{256a} being the preferred sequence in the $\epsilon \ll J\tau_p$ regime. As the contribution of finite-width pulse errors continues to grow, the first-order suppression in τ_p attained by $4 \times RGA_{64c}$ is more desirable; hence, the emergence of this optimal sequence when $\tau_p \gg \tau_d$ and $\epsilon \ll J\tau_p$.

In Table V, we present a summary of the performance scaling for each optimal sequence when subjected to finite-width pulses and flip-angle errors. The most interesting aspect is the eventual saturation in performance to terms linear in τ_p or quadratic in ϵ . This is an indication that attaining robustness for a wide range of τ_p , τ_d , and ϵ solely by manipulating the sequence configuration is insufficient. However, this does not invalidate the sequences we have obtained here since if a particular parameter regime is achievable, namely, the strong-pulse regime where flip-angle errors are dominant ($J\tau_p \ll J\tau_d \ll \epsilon$), sequences such as RGA_{8a} , RGA_{16a} and RGA_{64a} still exhibit a high level of robustness. This is not only evident from their ability to suppress errors solely proportional to ϵ and τ_d , but also from the fact that error accumulation for $\mathcal{O}(J\tau_p)$ terms is not observed as the number of pulses is increased; compare the effective error Hamiltonians of Eqs. (127), (132a)-(132c), and (140). Therefore, if the pulse duration can be made small relative to the pulse-interval then the finite-width errors are less consequential and it is still possible to maintain some form of robustness without the need for additional techniques. When such a regime is not attainable it may be necessary to utilize pulse shaping techniques to aid flip-angle error robust sequences in the suppression of finite-width pulses, or to exploit composite pulses to suppress flip-angle errors for sequences highly robust to finite-width pulses. Ultimately, a combination of sequence configuration, pulse shaping, and composite pulses is most likely the path forward to constructing fault-tolerant DD sequences for a wide range of parameter regimes.

Sequence	$\epsilon \ll J\tau_p \ll J\tau_d$	$J\tau_p \ll \epsilon \ll J\tau_d$	$J\tau_p \ll J\tau_d \ll \epsilon$	$\epsilon \ll J\tau_d \ll J\tau_p$	$J\tau_d \ll \epsilon \ll J\tau_p$	$J\tau_d \ll J\tau_p \ll \epsilon$
RGA_2	$\mathcal{O}(J\tau_c)$	$\mathcal{O}(J\tau_c)$	$\mathcal{O}(J\tau_c)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$
RGA_4	$\mathcal{O}(J^2\tau_d^2)$	$\mathcal{O}(J^2\tau_d^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(\epsilon^2)$
RGA_{8a}	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(\epsilon J\tau_d)$	$\mathcal{O}(\epsilon J\tau_d)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$
RGA_{8c}	$\mathcal{O}(J^2\tau_d^2)$	$\mathcal{O}(J^2\tau_d^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(J^2\tau_p\tau_d)$	$\mathcal{O}(\epsilon J\tau_p)$	$\mathcal{O}(\epsilon^2)$
RGA_{16a}	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$
$RGA_{16b'}$	$\mathcal{O}(\epsilon)$	$\mathcal{O}(\epsilon)$	$\mathcal{O}(\epsilon)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(\epsilon^2)$
RGA_{32a}	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(\epsilon^2)$
RGA_{64a}	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$
RGA_{64c}	$\mathcal{O}(\epsilon J\tau_p)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon J\tau_p)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(\epsilon^2)$
RGA_{256a}	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(\epsilon^2)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(J\tau_p)$	$\mathcal{O}(\epsilon^2)$

TABLE V. Summary of distance measure D scalings for optimal RGA_K sequences identified for DD evolution subjected to finite-width pulses of duration τ_p and flip-angle errors with rotation error ϵ in the regime of system-environment interaction-dominated ($J \gg \beta$) dynamics. The sequences with the best performance scaling in each parameter regime (column) are boxed. Note specifically the case of strong pulses dominated by flip-angle errors ($J\tau_p \ll J\tau_d \ll \epsilon$) where RGA_{8a} , RGA_{16a} , and RGA_{64a} obtain the more favorable performance scaling.

V. COMPARISON WITH KNOWN DETERMINISTIC SCHEMES

In this section, GA_K and RGA_K optimal sequences are compared to two known deterministic DD schemes: CDD and QDD, for each pulse profile.

CDD represents a fair comparison to the numerically optimal sequences since in both cases the pulse-intervals are fixed and the error suppression properties are dictated only by the sequence structure. Schemes which rely on optimized pulse delays to gain decoupling efficiency, such as QDD (see below), have much better scaling and error suppression properties than fixed delay schemes in the ideal pulse limit. However, the GA_K optimal sequences can be expected to prevail in the case of non-ideal pulse profiles, where no robust version of QDD currently exists. Optimal control theory can also be used to generate robust DD sequences [78].

The foundation of single-qubit CDD rests on the universal decoupling group (modulo irrelevant phases) $\mathcal{S} = \{I, X, Y, Z\}$, i.e., the single-qubit Pauli group, which implements first order error suppression by symmetrizing H_{err} over \mathcal{S} . Higher order error suppression can be achieved by continuing the symmetrization recursively, such that each additional level averages out the leading term in the Magnus expansion of each sub-level. At the r th level of recursion,

$$\text{CDD}_r = \bar{Y} \text{CDD}_{r-1} \bar{X} \text{CDD}_{r-1} \bar{Y} \text{CDD}_{r-1} X \text{CDD}_{r-1} \quad (144)$$

where $\text{CDD}_0 = f_{\tau_d}$ and the first level of symmetrization, $r = 1$, corresponds to RGA'_4 with $\{P_1, P_2\} = \{X, Y\}$:

$$\begin{aligned} \text{CDD}_1 &= RGA'_4 \\ &= (Y f_{\tau_d} Y)(Z f_{\tau_d} Z)(X f_{\tau_d} X)(I f_{\tau_d} I). \end{aligned} \quad (145)$$

Note that \mathcal{S} -based CDD requires 4^r pulses to accomplish r th order error suppression [23].

A large body of work now exists concerning DD sequences with non-uniform pulse-intervals, a technique which enables a drastic improvement over the exponential scaling of CDD.

Uhrig DD (UDD) is one such method, which applies DD pulses separated by unequal time intervals, at instants determined by a closed form expression originally developed in the context of the spin-boson model [34]. Using N control pulses, UDD suppresses the first N orders of the time-dependent perturbation theory expansion along the direction which does not commute with the pulses, provided the bath spectral density contains a sharp high-frequency cutoff [34, 35], or for generic bounded bath operators [36]. For an analysis of the scaling properties and corresponding distance measure performance of UDD see Ref. [38]. UDD can be extended to combat general single-qubit decoherence by nesting two anti-commuting UDD sequences. The resulting quadratic DD (QDD) scheme [39], achieves $\min(M_1, M_2)$ th order decoupling using $M_1 \times M_2$ pulses in the ideal pulse limit, where M_j , $j = 1, 2$, refers to the number of pulses in each nested UDD sequence. The decoupling efficiency of QDD has been extensively studied and confirmed numerically [40] and analytically [41]. Further generalizations of UDD include concatenated UDD (CUDD) [37] and nested UDD (NUDD) [42]. CUDD alleviates the restriction of single-channel decoherence suppression, however, still suffers from the exponential scaling of CDD. NUDD essentially follows the nesting scheme of QDD, while providing a generalization to multi-qubit and multi-level systems. As in the case of UDD and QDD, NUDD has been proven to apply to general decoherence models with bounded bath operators [43, 44]. For an analysis of the scaling properties and corresponding distance measure performance of QDD and NUDD see Ref. [79].

Most relevant to us, because it decouples a single qubit from a general system-bath interaction, is the QDD sequence, generated by nesting two UDD sequences as

$$\begin{aligned} \text{QDD}_{M_1, M_2} &= \Gamma_2^{M_2+1} \prod_{j=1}^{M_2+1} \Gamma_2 \text{UDD}_{M_1}^{\Gamma_1}(\lambda_j^{(M_2)} \tau_d) \\ &= \text{UDD}_{M_2}^{\Gamma_2}[\text{UDD}_{M_1}^{\Gamma_1}(\tau_d)], \end{aligned} \quad (146)$$

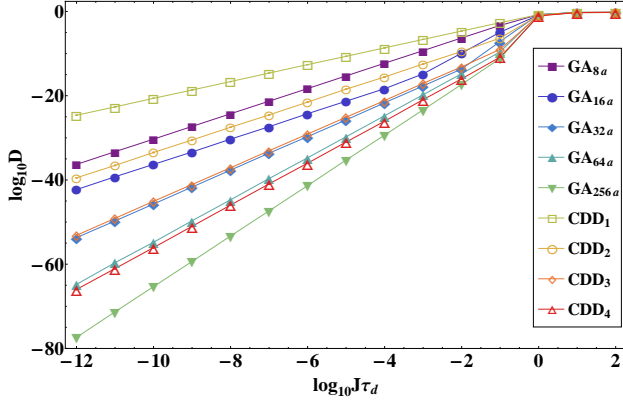


FIG. 7. (color online) Comparison of performance for CDD (empty symbols) and GA_K (filled symbols) as a function of the pulse-interval τ_d in the ideal pulse limit. The strength of the error Hamiltonian is chosen as $J = 1\text{MHz}$ and the strength of the pure-environment dynamics $\beta = 1\text{kHz}$. Optimal GA sequences achieve a higher order error suppression than CDD as the number of pulses increases; note GA_{64a} and GA_{256a} as compared to CDD_3 and CDD_4 , respectively.

where $\Gamma_1 \neq \Gamma_2$ are the generators of \mathcal{S} and

$$\text{UDD}_M^\Gamma(\tau_d) = \Gamma^{M+1} \prod_{k=1}^{M+1} \Gamma f_{\lambda_k \tau_d} \quad (147)$$

The free evolution periods for each UDD sequence are dictated by the normalized pulse-intervals

$$\lambda_k^{(M)} = \frac{t_k^{(M)} - t_{k-1}^{(M)}}{t_1^{(M)} - t_0^{(M)}}, \quad (148)$$

where

$$t_k^{(M)} = \tau_c \sin^2\left(\frac{k\pi}{2M+2}\right), \quad j = 1, 2, \dots, M+1, \quad (149)$$

and (minimum) pulse delay τ_d . In the following analysis, we will focus on $M_j = M$, $j = 1, 2$, to account for the effectively uniform decoherence model of Eq. (15) and arbitrarily choose $\Gamma_j \in \{Z, X\}$ as the generators of \mathcal{S} .

A. Ideal Pulses

In Figure 7, we compare the performance of CDD_l to GA_K with respect to $J\tau_d \in [10^{-12}, 10^2]$ for $J = 1\text{MHz}$ and $\beta = 1\text{kHz}$ in the ideal pulse limit. Numerically optimal sequences first coincide with CDD_l at $K = 16$, where both achieve second order error suppression. The main advantage of GA_{16a} over CDD_2 is a reduction in the error amplitude by a factor of approximately 10^3 . Significant improvement in error suppression is observed for $K = 64$ and $K = 256$, where GA optimal sequences offer an additional order of error suppression over corresponding CDD sequences, $r = 3, 4$. The results indicate that CDD does not constitute an optimal

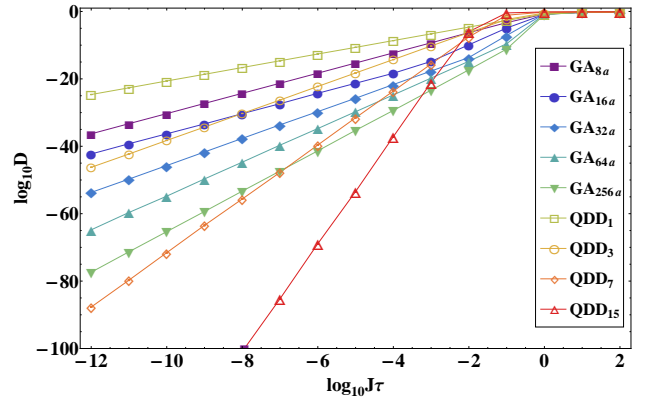


FIG. 8. (color online) Comparison of performance for QDD (empty symbols) and GA_K (filled symbols) as a function of the pulse-interval τ_d in the ideal pulse limit. The strength of the error Hamiltonian is chosen as $J = 1\text{MHz}$ and the strength of the pure-environment dynamics $\beta = 1\text{kHz}$. QDD outperforms GA_K for all sequence lengths, consistent with the expected superiority of interval-optimized schemes in the ideal pulse limit.

deterministic sequence structure for fixed pulse delays. In Sec. IV A 6, we elaborate on this fact and explore the possibility of designing a deterministic scheme to correctly describe the optimal GA sequences.

Pulse interval optimized sequences have been shown to far surpass the decoupling efficiency of any known fixed pulse-interval scheme in the ideal pulse limit, requiring only a quadratic increase in the number of pulses to suppress an additional order of the Magnus expansion. We validate the above statement by comparing QDD_M to the optimal GA sequences in Fig. 8 for $M = 1, 3, 7, 15$; an equivalent number of pulses to $K = 4, 16, 64, 256$, respectively. For each set of sequence orders $\{M, M\}$ QDD_M attains a decoupling order of M , which is beyond the ability of the GA_K for an equivalent number of pulses. QDD_M superiority is an indication that optimizing with respect to pulse-interval, in addition to pulse configuration, is necessary to obtain higher decoupling efficiency.

B. Finite-width Pulses

In Figs. 9 and 10, CDD_r and QDD_M , respectively, are compared to the RGA_K sequences optimized for finite pulse width: $K = 4, 8c, 16b, 32b, 64c, 256$. The pulse-interval is chosen as $\tau_d = 0.1\text{ns}$, while $\tau_p/\tau_d \in [10^{-5}, 10^7]$, $J = 1\text{MHz}$, and $\beta = 1\text{kHz}$. All results are averaged over 10 random instances of the Hamiltonian.

Optimal performance for RGA_K is observed predominately for $K = 32c, 64c$ with $K = 16a'$ exhibiting a more favorable performance only in the strong-pulse regime, namely when $\tau_p/\tau_d < 10^{-2}$ in Fig. 9. As the finite-width pulse errors contribute more substantially, $\mathcal{O}(J^2\tau_p^2)$ terms remaining in the effective error Hamiltonian for $K = 16c$ results in a rapid decrease in performance leading to RGA_{32c}/RGA_{64c}

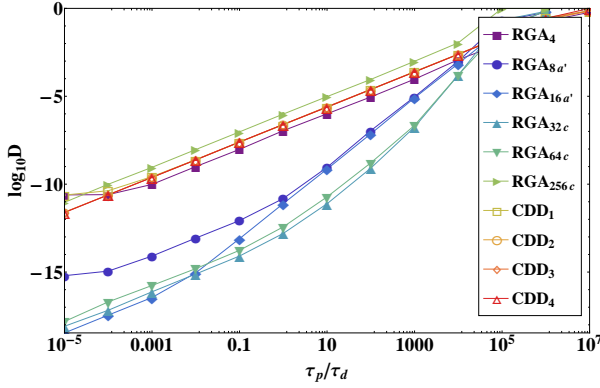


FIG. 9. (color online) Comparison of performance between RGA_K and CDD_r when subjected to finite pulse duration. Performance is characterized as a function of τ_p , while $\tau_d = 0.1\text{ns}$. CDD_r performance is essentially the same for all r , scaling as $D \sim \mathcal{O}(J\tau_p)$. RGA_K achieves a significant increase in robustness over CDD_r at $K = 8a', 16a', 32c, 64c$, where the performance surpasses the linear scaling in τ_p . Note the eventual saturation in decoupling order characteristic of the rectangular pulse profile displayed by the nearly equivalent scaling of $K = 32c, 64c$. All results are averaged over 10 realizations of B_μ . Error bars are included, but are quite small.

dominance. $K = 4$ maintains the lowest performance of all RGA_K sequences due to its inability to suppress the first order contribution in τ_p , $D \sim \mathcal{O}(J\tau_p)$. CDD_r performance is nearly equivalent to $K = 4$ for $\tau_p/\tau_d \geq 10^{-4}$, scaling as $D \sim \mathcal{O}(J\tau_p)$ for all r . The most noticeable difference occurs at $\tau_p/\tau_d < 10^{-4}$, where CDD_r maintains the linear scaling in τ_p for $r = 2, 3, 4$ and surpasses $K = 4, 8c$. In Ref. [23], an analysis of CDD_r in the presence of finite pulse width is discussed as well. There it was shown that CDD_r can reduce pulse-width errors as the concatenation level increases if $\tau_p \ll \tau_d$. Although the total cycle time was fixed, as opposed to the pulse-interval, the results obtained here are quite similar in the $\tau_p/\tau_d < 10^{-4}$ regime and confirm the inherent robustness of CDD_r to finite-width pulse errors.

As discussed in Refs. [80], UDD-based schemes are quite susceptible to finite pulse-width errors and must be implemented with specially tailored pulses to regain a portion of the UDD decoupling efficiency. We confirm this result here for the most simplistic pulse shape: the rectangular pulse. Increasing the sequence order does not result in an increase, or sustainability, of performance; rather an accumulation of error results. As in the case of CDD_r , the performance maintains a linear scaling in τ_p throughout the specified range for the higher of the three sequence orders: $M = 3, 7, 15$. The only variation occurs at $M = 1$ where the performance becomes dependent upon τ_d for $\tau_p \ll \tau_d$. Although there exists a regime where both deterministic schemes outperform $K = 4, 8c$, higher order RGA_K sequences provide a level of robustness that cannot be matched by either CDD_r or QDD_M .

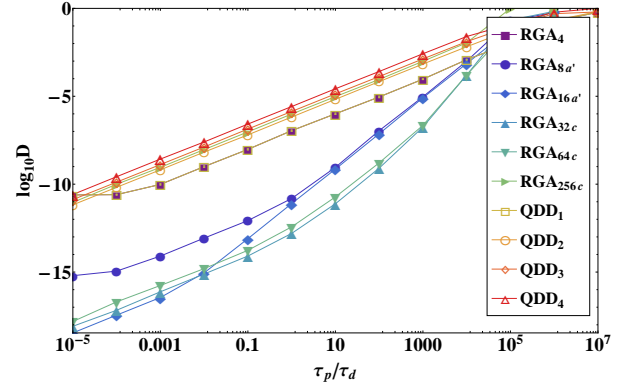


FIG. 10. (color online) Comparison of performance between RGA_K and QDD_M , $M = 1, 3, 7, 15$, when subjected to finite pulse duration. The pulse-interval is fixed at $\tau_d = 0.1\text{ns}$ and performance is analyzed as a function of τ_p . QDD_M performance becomes increasingly worse as the sequence order increases due to an accumulation in errors brought about by the finite duration of the pulses. As in the case of CDD_r , QDD_M performance maintains $D \sim \mathcal{O}(J\tau_p)$ for all M . See Fig. 9 for details regarding RGA_K performance.

C. Flip-angle errors

In Sec. IV C, robust sequences were identified for control pulses subjected to flip-angle errors. Here, we compare the numerically optimal RGA_K sequences to CDD_r and QDD_M as a function of ϵ . We consider the case of interaction-dominated dynamics and set the strengths of the environment dynamics and system-bath interaction to $\beta = 1\text{kHz}$ and $J = 1\text{MHz}$, respectively. The pulse delay is chosen as $\tau_d = 0.1\text{ns}$ and all data is averaged over 20 realizations of the bath operators $\{B_\mu\}$. In addition to $K = 2, 8a, 16a, 64a$, which exhibit an increase in decoupling order for terms proportional to ϵ , $K = 4, 32a, 256a$ are included in the comparison as well to fully characterize RGA_K performance with respect to both deterministic schemes.

We first focus on RGA_K and CDD_r , where RGA_K superiority is clearly evident for all values of K shown in Fig. 11. Optimal performance is observed for $K = 2, 8a, 16a, 64a$, as expected, with $K = 16a$ providing the highest level of robustness to flip-angle errors using the smallest number of pulses. Although the lowest level of performance for RGA_K occurs at $K = 4, 32a, 256a$, a considerable improvement over the corresponding CDD_r , $r = 1, 2, 3, 4$, exists. Optimal CDD_r performance is achieved at $r = 2, 4$, where the performance can be shown to scale as $D \sim \mathcal{O}(\epsilon^2)$. The remaining levels of concatenation, $r = 1, 3$ do not achieve first order suppression, therefore, $D \sim \mathcal{O}(\epsilon)$. The comparison clearly indicates that the numerically optimized sequences are highly robust to flip-angle errors and capable of achieving performances far beyond those of CDD_r .

Analyzing QDD_M , for $M = 1, 3, 7, 15$, as a function of ϵ , we find that the performance maintains a $D \sim \mathcal{O}(\epsilon^2)$ scaling for all M ; see Fig. 12. A similar performance scaling is obtained for $r = 2, 4$ in CDD_r as well, where the performance

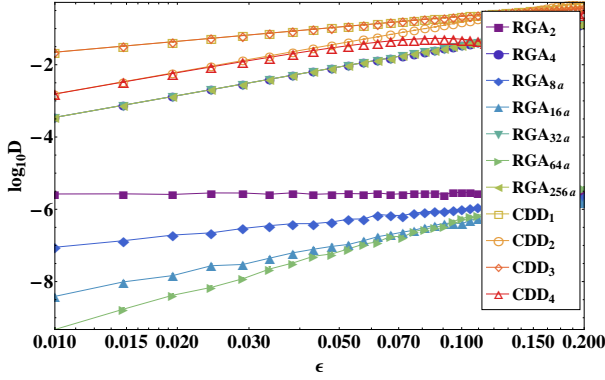


FIG. 11. (color online) Performance of RGA_K sequences versus CDD_l as a function of flip-angle error $\epsilon \in [0.01, 0.2]$ averaged over 20 realizations of B_μ . The relevant parameters are chosen as $J = 1\text{MHz}$, $\beta = 1\text{kHz}$, and $\tau_d = 0.1\text{ns}$. Numerically optimal sequences are found to be highly robust against flip-angle errors, significantly out-performing CDD_r for $r = 1, 2, 3, 4$.

is equivalent for both concatenation levels. In contrast, the performance for QDD_M diminishes with increasing M , indicating an accumulation of error rather than a reduction, or sustainability as in the case of CDD_r . Although variations in performance exist between the deterministic schemes, their robustness to flip-angle errors is clearly not comparable to the optimized RGA_K sequences.

D. Finite-width and Flip-angle Errors

As a final comparison, we examine the performance and robustness of RGA_K , CDD_r , and QDD_M in the presence of both finite-width and flip-angle pulse errors. In particular, we focus on the performance as a function of ϵ and τ_p relative to the size of the pulse-interval $\tau_d = 0.1\text{ns}$. The flip-angle error is varied from zero to a 15% rotation error, the ratio of pulse duration to interpulse delay $\tau_p/\tau_d \in [10^{-5}, 10^4]$, and $J = 1\text{MHz}$ with $\beta = 1\text{kHz}$. We average over 10 realizations of B_μ .

The performance of RGA_K is shown in Fig. 13 in panels (a)-(d) for $K = 4, 16a, 64a, 256a$, respectively. Although additional optimal configurations were identified at other sequence lengths, we focus on these particular values of K to compare them directly to CDD_r , $r = 1, 2, 3, 4$, and QDD_M , $M = 1, 3, 7, 15$. RGA_K performance is found to be primarily dependent upon τ_p , only showing significant ϵ -dependence when $\epsilon \leq 0.04$. Below $\epsilon = 0.04$, specifically in the region where $\tau_p < \tau_d$, robustness increases as the number of pulses is increased from $K = 16a$ to $K = 256a$. Error accumulation within this particular range of values appears to not be an issue.

CDD_r performance is displayed in panels (e)-(h) of Fig. 13 for $r = 1, 2, 3, 4$, respectively. Contrary to RGA_K , CDD_r performance exhibits a strong dependence on flip-angle errors rather than finite pulse duration. Optimal performance is

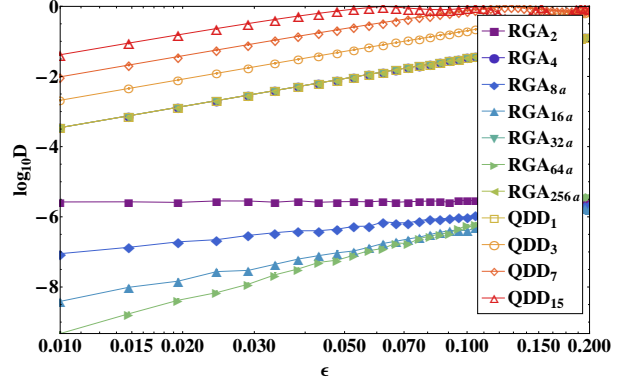


FIG. 12. (color online) Performance of RGA_K sequences versus QDD_M as a function of flip-angle error $\epsilon \in [0.01, 0.2]$ averaged over 10 realizations of B_μ . The sequence orders are chosen as $M = 3, 7, 15$ and directly correspond to $K = 16, 64, 256$. QDD is shown to be highly sensitive to flip-angle errors, decreasing in performance as M grows. Again, robust GA sequences achieve optimal performance. See Fig. 11 for additional details regarding simulation parameters.

heavily concentrated around small values of ϵ due to the low level of robustness exhibited by CDD_r for flip-angle errors. Robustness to finite pulse duration appears to be most noticeable for $r = 2, 4$, although performance is still rather poor compared to RGA_K .

Lastly, we examine QDD_M in panels (i)-(l) in Fig. 13, where $M = 1, 3, 7, 15$, respectively. The lowest sequence order, $M = 1$, generates the exact same sequence as RGA_4 , therefore, performance is identical. The remaining sequence orders result in continual error accumulation, which is evident from the steady decline in performance from $M = 3$ to $M = 15$. As in the case of CDD_r , QDD_M performance is primarily ϵ -dependent and degrades rapidly with increasing ϵ . Similarly to the additional comparisons of numerically optimized sequences and deterministic schemes for faulty DD pulses given above, RGA_K sequences significantly outperform CDD_r and QDD_M .

VI. EXISTENCE OF DETERMINISTIC STRUCTURE

Concatenation appears to play an important role in the construction of optimal DD in the limit of fixed pulse-intervals. One of the goals of this study is to determine whether the optimal sequences provided above can be generalized into a deterministic concatenation scheme for arbitrary order decoupling. As suggested by the analysis summarized in Table II, we find that such a scheme is possible by utilizing GA_{8a} as the fundamental unit of concatenation in

$$GA_{8a}^{(q)} = GA_{8a}[GA_{8a}^{(q-1)}], \quad (150)$$

where $GA_{8a}^{(0)} \equiv f_{\tau_d}$. Requiring 8^q pulses, $GA_{8a}^{(q)}$ achieves $2q$ th order error suppression; a factor of 2^q improvement over the decoupling efficiency of CDD , which demands 4^{2q} pulses

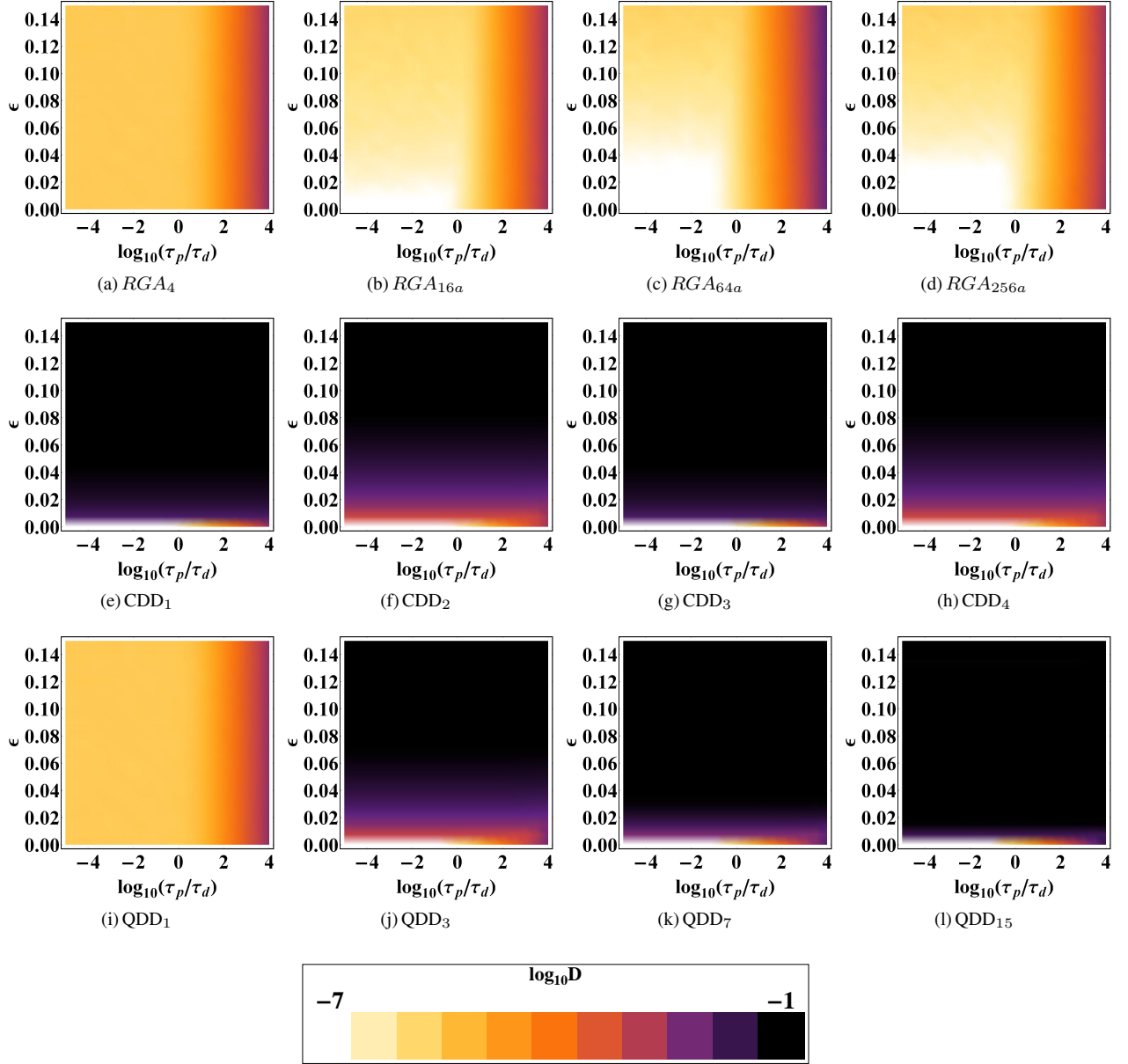


FIG. 13. (color online) Performance of RGA_K , CDD_r , and QDD_M when subjected to finite pulse duration and flip-angle errors. The pulse-interval is fixed at $\tau_d = 0.1\text{ns}$, while $J = 1\text{MHz}$ and $\beta = 1\text{kHz}$. RGA_K sequences significantly outperform both CDD_r and QDD_M for $K = 4, 16, 64, 256$. The most notable region of robustness exists for $\epsilon < 0.04$ and $\tau_p < \tau_d$ for $K = 16, 64, 256$.

to achieve an equivalent decoupling order. The increased decoupling efficiency is facilitated by second order error suppression provided by GA_{8a} , which essentially boosts the efficiency by a factor of 2 at each level of concatenation.

In Fig. 14, we compare the performance of $GA_{8a}^{(q)}$ to that of CDD_r and GA_4 in the case of zero-width pulses as a function of the number of pulses K . The results are averaged over 25 random realizations of B_μ , where the pulse operators are designated by $\{P_1, P_2\} = \{X, Y\}$ for each generalized sequences. The strengths of the error Hamiltonian and bath dynamics are given by $J = 1\text{MHz}$ and $\beta = 1\text{kHz}$, respectively, and the minimum pulse delay $\tau_d = 0.1\text{ns}$. In com-

paring $q = 1, 2, \dots, 4$ and $l = 1, 2, \dots, 6$, the performance of $GA_{8a}^{(q)}$ improves dramatically as the level of concatenation increases, far exceeding that of CDD_r and GA_4 .

However, the truly meaningful test of sequence performance is in the presence of pulse errors. Defining

$$RGA_{8a}^{(q)} = RGA_{8a}[RGA_{8a}^{(q-1)}], \quad (151)$$

to effectively combat the inclusion of flip-angle and finite-width pulse errors, we compare the performance of $RGA_{8a}^{(q)}$ to CDD_r and RGA_4 in Fig. 15 for $\{P_1, P_2\} = \{Y, X\}$. The relevant Hamiltonian parameters are equivalent to those cho-

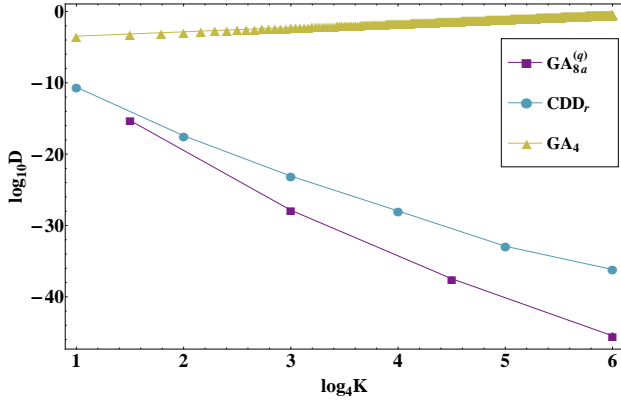


FIG. 14. (color online) Comparison of $GA_{8a}^{(q)}$, CDD_r , and XY_4 performance after one cycle as a function of the number of pulses K for ideal, zero-width pulses. The strength of the error Hamiltonian and environment dynamics are set to $J = 1\text{MHz}$ and $\beta = 1\text{kHz}$, respectively, and the minimum pulse-interval $\tau_d = 0.1\text{ns}$. Results are averaged over 25 random realizations of B_μ , where the error bars are shown, but quite small. As expected by the results of the GA search $GA_{8a}^{(q)}$, $q = 1, \dots, 4$ is indeed superior to CDD_r , $r = 1, \dots, 6$, and XY_4 .

sen for the ideal case, while the flip-angle error $\epsilon = 0.01$. As opposed to fixing τ_d , we select a fixed cycle time τ_c to analyze the relationship between τ_p and τ_d as the number of pulses grows. In particular, we consider $\tau_c = 1\text{ns}$ and $\tau_p/\tau_c = 10^{-10}$. We expect robustness to be most noticeable in the strong pulse regime, $\tau_p \gg \tau_d$, where the primary form of pulse error is due to the flip-angle errors. Performing the analysis with a fixed cycle time allows us to examine robustness as a function of concatenation level and as $\tau_p \rightarrow \tau_d$, simultaneously. As expected by direct calculation of the effective Hamiltonian for RGA_4 [see Eq. (124)], increasing the number of pulses via multiple DD cycles does not offer an enhancement in sequence performance due to an immediate accumulation of error proportional to τ_p and ϵ^2 . In contrast, CDD_r performance remains fairly consistent and oscillates between two values, seemingly dependent on the parity of the concatenation level. In Refs. [23, 25], a similar study of CDD performance conveyed that effective cancellation of pulse-width errors occurs for a range of concatenation levels if the pulse width is much smaller than the pulse-interval; eventual saturation in performance as r increases occurs once this condition is violated. The performance characteristics of CDD_r differ here due to the presence of flip-angle errors, which are not accounted for in Refs. [23, 25]. Combined errors are most effectively addressed by $RGA_{8a}^{(q)}$, as can be seen by the improvements in sequence performance as concatenation level increases. Note that the performance eventually begins to show signs of saturation as the concatenation level increases. Essentially, the pulse-interval is approaching a value comparable to the pulse duration which leads to finite-width errors becoming a more significant decoherence mechanism than flip-angle errors or the error Hamiltonian. RGA_{8a} and its concatenated versions do not provide protection against finite-width

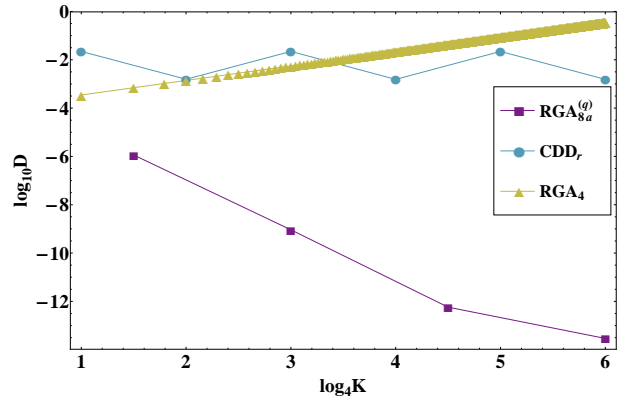


FIG. 15. (color online) Comparison of $RGA_{8a}^{(q)}$, CDD_r , and RGA_4 performance versus the number of pulses K for combined pulse errors (flip-angle and finite-width) after one cycle. Hamiltonian parameters, J and β , are the same as those in Fig. 14 with the total cycle time fixed at $\tau_c = 1\text{ns}$, as opposed to τ_d . Results are averaged over 25 realizations of B_μ with $\epsilon = 0.01$ and $\tau_p/\tau_c = 10^{-10}$. The performance of $RGA_{8a}^{(q)}$ improves as the number of pulses is increased within a given cycle time τ_c . For the specified parameters, CDD_r does not exhibit enhanced performance with increasing concatenation level.

errors, and therefore their performance becomes hindered by the presence of terms that are first order in τ_p .

VII. CONCLUSIONS

In this work we showed that numerically optimal DD sequences can be constructed using a genetic algorithm in conjunction with a simulated annealing convergence accelerator and a novel complexity-reduction technique. The search focused on sequences containing $K = 1, 2, \dots, 256$ pulses, however, we identified optimal performance at $K = 4, 8, 16, 32, 64, 256$ and compared each sequence to known deterministic schemes, such as CDD and QDD. The ideal-pulse analysis showed that, under the constraint of a fixed pulse interval, optimal sequences can be constructed which outperform CDD, yet fall short of the decoupling efficiency realized by QDD. Optimization proved to be quite beneficial in the case of finite-width and flip-angle errors, where numerically optimal sequences obtained a level of robustness that could not be reached by either CDD or QDD.

By analyzing sequence structure, we found that concatenation appears to be the preferred method for high order decoupling for all pulse profiles considered. This technique is most beneficial when each concatenation sub-level attains the same decoupling order. As most notably observed for faulty DD pulses subjected to flip-angle errors [$K = 32, 256$], concatenating sequences with unequal decoupling orders results in considerable attenuations in sequence performance and error suppression that can be attributed to the inherent non-robustness of the low-order decoupling sequence.

The culmination of our study is centered around the identification of a deterministic sequence structure that obtains a

high decoupling efficiency in the ideal limit and robustness to errors generated by pulse imperfections. We determined that RGA_{8a} [Eq. (105)] is the favored generating sequence for a majority of the pulse profiles considered. Concatenating this particular sequence, it is possible to suppress the first $2q$ terms in the Magnus expansion using 8^q pulses. Compared to the 4^q pulses required for the widely used original version of CDD [22], the concatenated version of RGA_{8a} utilizes 2^q fewer pulses to obtain the same decoupling order. Although it is not possible to obtain such high levels of decoupling in the presence of pulse imperfections, RGA_{8a} contains an inherent robustness that continues to aid the error suppression process as the level of concatenation increases. This result is most apparent in our final study of faulty DD pulses, which includes both finite duration and flip-angle errors; we found the concatenated RGA_{8a} construction to be the most robust scheme available for fixed pulse-interval DD sequences.

The importance of optimizing over pulse-intervals was clearly displayed in the ideal pulse analysis, where sequence configuration optimization alone could not supply the decoupling efficiency achieved by QDD. However, in agreement with previous work, we have shown that pulse-interval optimized sequences fail to be robust against additional errors generated by faulty DD pulses. Future work should focus on

extending the search algorithm to incorporate multi-qubit systems and unequal pulse delays, to obtain robust DD sequences in the presence of various pulse errors.

ACKNOWLEDGMENTS

We thank Dr. Joaquin Cerda Boluda for his valuable contribution towards constructing a highly efficient GA algorithm. We would also like to thank Gerardo Paz Silva, Gonzalo A. Alvarez, Dieter Suter, and Wan-Jung Kuo for useful discussions and valuable input. This research was supported by the ARO MURI grant W911NF-11-1-0268, by the Department of Defense, by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20165, and by NSF grants No. CHE-924318 and CHE-1037992. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

-
- [1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2000).
 - [2] L. Viola and S. Lloyd, *Phys. Rev. A* **58**, 2733 (1998).
 - [3] L. Viola, E. Knill, and S. Lloyd, *Phys. Rev. Lett.* **82**, 2417 (1999).
 - [4] P. Zanardi, *Phys. Lett. A* **258**, 77 (1999).
 - [5] D. Vitali and P. Tombesi, *Phys. Rev. A* **59**, 4178 (1999).
 - [6] L.-M. Duan and G.-C. Guo, *Phys. Lett. A* **261**, 139 (1999).
 - [7] M. Byrd and D. Lidar, *Quantum Inf. Proc.* **1**, 19 (2002).
 - [8] D. J. Szwer, S. C. Webster, A. M. Steane, and D. M. Lucas, *ArXiv* (2012), [quant-ph/1009.6189](#).
 - [9] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, *Phys. Rev. A* **79**, 062324 (2009).
 - [10] A. Bermudez, P. O. Schmidt, M. B. Plenio, and A. Retzker, *ArXiv* (2012), [quant-ph/1110.1870](#).
 - [11] A. Ajoy, G. A. Álvarez, and D. Suter, *Phys. Rev. A* **83**, 032303 (2011).
 - [12] A. M. Souza, G. A. Álvarez, and D. Suter, *Phys. Rev. Lett.* **106**, 240501 (2011).
 - [13] A. M. Souza, G. A. Álvarez, and D. Suter, *Phys. Rev. A* **85**, 032306 (2012).
 - [14] C. Barthel, J. Medford, C. M. Marcus, M. P. Hanson, and A. C. Gossard, *Phys. Rev. Lett.* **105**, 266808 (2010).
 - [15] A. Laraoui and C. A. Meriles, *Phys. Rev. B* **84**, 161403 (2011).
 - [16] B. Naydenov, F. Dolde, L. T. Hall, C. Shin, H. Fedder, L. C. L. Hollenberg, F. Jelezko, and J. Wrachtrup, *Phys. Rev. B* **83**, 081201 (2011).
 - [17] Z. Wang, G. de Lange, D. Riste, R. Hanson, and V. V. Dobrovitski, *ArXiv* (2012), [quant-ph/1202.0462](#).
 - [18] E. L. Hahn, *Phys. Rev.* **80**, 580 (1950).
 - [19] S. Meiboom and D. Gill, *Rev. Sci. Instrum.*, 688 (1958).
 - [20] A. A. and Maudsley, *Journal of Magnetic Resonance* (1969) **69**, 488 (1986).
 - [21] T. Gullion, D. B. Baker, and M. S. Conradi, *Journal of Magnetic Resonance* (1969) **89**, 479 (1990).
 - [22] K. Khodjasteh and D. A. Lidar, *Phys. Rev. Lett.* **95**, 180501 (2005).
 - [23] K. Khodjasteh and D. A. Lidar, *Phys. Rev. A* **75**, 062310 (2007).
 - [24] D. A. Lidar, P. Zanardi, and K. Khodjasteh, *Phys. Rev. A* **78**, 012308 (2008).
 - [25] H. K. Ng, D. A. Lidar, and J. Preskill, *Phys. Rev. A* **84**, 012305 (2011).
 - [26] W. M. Witzel and S. Das Sarma, *Phys. Rev. B* **76**, 241303 (2007).
 - [27] W. Zhang, V. V. Dobrovitski, L. F. Santos, L. Viola, and B. N. Harmon, *Phys. Rev. B* **75**, 201302 (2007).
 - [28] W. Zhang, N. P. Konstantinidis, V. V. Dobrovitski, B. N. Harmon, L. F. Santos, and L. Viola, *Phys. Rev. B* **77**, 125336 (2008).
 - [29] X. Peng, D. Suter, and D. A. Lidar, *J. Phys. B* **44**, 154003 (2011).
 - [30] G. A. Álvarez, A. Ajoy, X. Peng, and D. Suter, *Phys. Rev. A* **82**, 042306 (2010).
 - [31] Z.-H. Wang, W. Zhang, A. M. Tyryshkin, S. A. Lyon, J. W. Ager, E. E. Haller, and V. V. Dobrovitski, *Phys. Rev. B* **85**, 085206 (2012).
 - [32] A. M. Tyryshkin, Z. Wang, W. Zhang, E. E. Haller, J. W. Ager, V. V. Dobrovitski, and S. A. Lyon, (2010), eprint [arXiv:1011.1903](#).
 - [33] G. A. Álvarez, A. M. Souza, and D. Suter, *Phys. Rev. A* **85**, 052324 (2012).
 - [34] G. S. Uhrig, *Phys. Rev. Lett.* **98**, 100504 (2007).
 - [35] S. Pasini and G. S. Uhrig, *Phys. Rev. A* **81**, 012309 (2010).

- [36] W. Yang and R.-B. Liu, *Phys. Rev. Lett.* **101**, 180403 (2008).
- [37] G. S. Uhrig, *Phys. Rev. Lett.* **102**, 120502 (2009).
- [38] G. S. Uhrig and D. A. Lidar, *Phys. Rev. A* **82**, 012301 (2010).
- [39] J. R. West, B. H. Fong, and D. A. Lidar, *Phys. Rev. Lett.* **104**, 130501 (2010).
- [40] G. Quiroz and D. A. Lidar, *Phys. Rev. A* **84**, 042328 (2011).
- [41] W.-J. Kuo and D. A. Lidar, *Phys. Rev. A* **84**, 042329 (2011).
- [42] Z.-Y. Wang and R.-B. Liu, *Phys. Rev. A* **83**, 022306 (2011).
- [43] L. Jiang and A. Imambekov, *Phys. Rev. A* **84**, 060302 (2011).
- [44] W.-J. Kuo, G. A. Paz-Silva, G. Quiroz, and D. A. Lidar, “Universality proof and analysis of generalized nested Uhrig dynamical decoupling,” (2012), eprint arXiv:1207.1665.
- [45] L. Viola and E. Knill, *Phys. Rev. Lett.* **90**, 037901 (2003).
- [46] K. Khodjasteh and L. Viola, *Phys. Rev. Lett.* **102**, 080501 (2009).
- [47] K. Khodjasteh and L. Viola, *Phys. Rev. A* **80**, 032314 (2009).
- [48] K. Khodjasteh, D. A. Lidar, and L. Viola, *Phys. Rev. Lett.* **104**, 090501 (2010).
- [49] Z. Xiao, L. He, and W.-g. Wang, *Phys. Rev. A* **83**, 032322 (2011).
- [50] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, *Nature* **458**, 996 (2009).
- [51] K. Khodjasteh, T. Erdélyi, and L. Viola, *Phys. Rev. A* **83**, 020305 (2011).
- [52] H. Uys, M. J. Biercuk, and J. J. Bollinger, *Phys. Rev. Lett.* **103**, 040501 (2009).
- [53] D. Hayes, K. Khodjasteh, L. Viola, and M. J. Biercuk, *ArXiv* (2011), quant-ph/1109.6002.
- [54] A. Souza, G. Alvarez, and D. Suter, *ArXiv* (2011), quant-ph/1110.6334.
- [55] L. Viola, S. Lloyd, and E. Knill, *Phys. Rev. Lett.* **83**, 4888 (1999).
- [56] K. Khodjasteh and D. A. Lidar, *Phys. Rev. A* **78**, 012355 (2008).
- [57] J. R. West, D. A. Lidar, B. H. Fong, and M. F. Gyure, *Phys. Rev. Lett.* **105**, 230503 (2010).
- [58] L.-A. Wu and D. A. Lidar, *Phys. Rev. Lett.* **88**, 207902 (2002).
- [59] M. S. Byrd and D. A. Lidar, *Phys. Rev. Lett.* **89**, 047901 (2002).
- [60] U. Haeberlen, *High Resolution NMR in Solids, Advances in Magnetic Resonance Series Supplement 1* (Academic Press, New York, 1976).
- [61] L. Viola and E. Knill, *Phys. Rev. Lett.* **90**, 037901 (2003).
- [62] W. Magnus, *Communications on Pure and Applied Mathematics* **7**, 649 (1954).
- [63] S. Blanes, F. Casas, J. Oteo, and J. Ros, *Phys. Rep.* **470**, 151 (2009).
- [64] P. C. Moan, J. A. Oteo, and J. Ros, *Journal of Physics A: Mathematical and General* **32**, 5133 (1999).
- [65] R. Ernst, G. Bodenhausen, and A. Wokaum, *Principles of nuclear magnetic resonance in one and two dimensions* (Clarendon Press, United Kingdom, 1987).
- [66] Z. Wang, W. Zhang, A. M. Tyryshkin, S. A. Lyon, J. W. Ager, E. E. Haller, and V. V. Dobrovitski, *ArXiv* (2010), quant-ph/1011.6417.
- [67] A. M. Tyryshkin, Z. H. Wang, W. Zhang, E. E. Haller, J. W. Ager, V. V. Dobrovitski, and S. A. Lyon, *ArXiv* (2010), quant-ph/1011.1903.
- [68] Z.-H. Wang and V. V. Dobrovitski, *ArXiv* (2010), quant-ph/1101.0292.
- [69] R. L. Kosut, M. Grace, C. Brif, and H. Rabitz, *ArXiv* (2006), quant-ph/0606064.
- [70] M. D. Grace, J. Dominy, R. L. Kosut, C. Brif, and H. Rabitz, *New J. of Phys.* **12**, 015001 (2010).
- [71] A. Petiteau, Y. Shang, S. Babak, and F. Feroz, *Phys. Rev. D* **81**, 104016 (2010).
- [72] L. P. Pryadko and G. Quiroz, *Phys. Rev. A* **77**, 012330 (2008).
- [73] L. P. Pryadko and G. Quiroz, *Phys. Rev. A* **80**, 042317 (2009).
- [74] S. Pasini, P. Karbach, C. Raas, and G. S. Uhrig, *Phys. Rev. A* **80**, 022328 (2009).
- [75] S. Pasini, T. Fischer, P. Karbach, and G. S. Uhrig, *Phys. Rev. A* **77**, 032315 (2008).
- [76] S. Pasini and G. S. Uhrig, *Journal of Physics A: Mathematical and Theoretical* **41**, 312005 (2008).
- [77] C. H. Wang and J. D. Ramshaw, *Phys. Rev. B* **6**, 3253 (1972).
- [78] Y. Tabuchi and M. Kitagawa, *ArXiv* (2012), 1208.5218.
- [79] Y. Xia, G. S. Uhrig, and D. A. Lidar, *Phys. Rev. A* **84**, 062332 (2011).
- [80] S. Pasini, P. Karbach, and G. S. Uhrig, *ArXiv* (2011), 1009.2638.

Appendix A: Scaling of Performance

Here, we prove the scaling of distance measure $D(U, I_S)$ described in Eq. (26) for the ideal pulse limit. First, we prove that the general distance measure $D(U, G)$ [see Eq. (22)] satisfies

$$D(U, G) \leq \frac{1}{\sqrt{2}} \|U - G \otimes I_B\|. \quad (\text{A1})$$

The upper-bound is obtained by utilizing the steps originally taken in Ref. [69] to obtain the closed form expression of $D(U, G)$ given in Eq. (23), where initially it is shown that, for Φ satisfying $\Phi^\dagger \Phi = I_B$,

$$\begin{aligned} D(U, G) &= \frac{1}{\sqrt{2} d_S d_B} \min_{\Phi} \|U - G \otimes \Phi\|_F \\ &= \min_{\Phi} \sqrt{1 - \frac{1}{d_S d_B} \text{Re}\{\text{Tr}[U(G^\dagger \otimes \Phi^\dagger)]\}} \\ &= \min_{\Phi} \sqrt{1 - \frac{1}{d_S d_B} \text{Re}\{\text{Tr}_B[\text{Tr}_S[U(G^\dagger \otimes I_B)]\Phi^\dagger]\}} \\ &= \min_{\Phi} \sqrt{1 - \frac{1}{d_S d_B} \text{Re}\{\text{Tr}(\Gamma \Phi^\dagger)\}}. \end{aligned} \quad (\text{A2})$$

It is then noticed that computing the minimization problem of Eq. (22) is equivalent to finding the maximum value of $\text{Re}\{\text{Tr}(\Gamma \Phi^\dagger)\}$ over all unitary Φ . In order to complete the proof, the singular value decomposition (SVD) $\Gamma = W \Sigma V^\dagger$ is invoked, where W, V are unitary and $\Sigma = \text{diag}(s_1, \dots, s_{d_B})$ is a real diagonal matrix containing the singular values $s_1 \geq s_2 \geq \dots \geq s_{d_B} \geq 0$. The relevant expression becomes

$$\text{Tr}(\Gamma \Phi^\dagger) = \text{Tr}(W \Sigma V^\dagger \Phi^\dagger) = \text{Tr}[\Sigma(V^\dagger \Phi^\dagger W)] \quad (\text{A3})$$

and the final details of the proof involve showing that $\text{Re}\{\text{Tr}[\Sigma(V^\dagger \Phi^\dagger W)]\}$ is essentially maximized if and only if $V^\dagger \Phi^\dagger W = I_B$, or equivalently when $\Phi = W V^\dagger$. Note that

$$\text{Re}\{\text{Tr}[\Sigma(V^\dagger \Phi^\dagger W)]\} \leq \text{Tr}(\Sigma) \quad (\text{A4})$$

holds for all Φ that differ from $W V^\dagger$.

It is now straightforward to obtain the bound expressed in Eq. (A1) since by choosing $\Phi = I_B$ we are satisfying the

lower bound expression of Eq. (A4) and generating the upper-bound

$$D(U, G) \leq \frac{1}{\sqrt{2d_S d_B}} \|U - G \otimes I_B\|_F \quad (\text{A5})$$

$$\leq \frac{1}{\sqrt{2}} \|U - G \otimes I_B\|. \quad (\text{A6})$$

Note that we have used the inequality $\|T\|_F \leq \sqrt{d_S d_B} \|T\|$ in order to transform from the Frobenius norm [Eq. (A5)] to the sup-operator norm [Eq. (A6)].

Choosing $G = I_S$, as specified by the desired action of the DD evolution on the system, and the toggling frame evolution operator $U = \tilde{U}_0(\tau_c)$, the distance measure upper-bound becomes

$$D \leq \frac{1}{\sqrt{2}} \|\tilde{U}_0(\tau_c) - I_S \otimes I_B\|. \quad (\text{A7})$$

Expressing the unitary evolution operator $\tilde{U}_0(\tau_c)$ as a time-dependent perturbation expansion

$$\tilde{U}_0(\tau_c) = I_S \otimes I_B + \sum_{n=1}^{\infty} \tilde{U}_0^{(n)}(\tau_c), \quad (\text{A8})$$

with Dyson operators

$$\tilde{U}_0^{(n)}(\tau_c) = (-i)^n \int_0^{\tau_c} dt_1 \cdots \int_0^{t_{n-1}} dt_n \prod_{j=1}^n \tilde{H}_0(t_j), \quad (\text{A9})$$

we obtain, using the triangle inequality,

$$D \leq \frac{1}{\sqrt{2}} \left\| \sum_{n=1}^{\infty} \tilde{U}_0^{(n)}(\tau_c) \right\| \leq \frac{1}{\sqrt{2}} \sum_{n=1}^{\infty} \|\tilde{U}_0^{(n)}(\tau_c)\|. \quad (\text{A10})$$

The final step is to show that $\|\tilde{U}_0^{(n)}(\tau_c)\|$ obtains the scaling claimed by Eq. (26). This is accomplished by using (1) the triangle inequality, (2) sub-multiplicativity, and (3) unitary invariance to obtain an upper-bound on Eq. (A9) as follows:

$$\begin{aligned} \|\tilde{U}_0^{(n)}(\tau_c)\| &= \|(-i)^n \int_0^{\tau_c} dt_1 \cdots \int_0^{t_{n-1}} dt_n \prod_{j=1}^n \tilde{H}_0(t_j)\| \\ &\stackrel{(1)}{\leq} \int_0^{\tau_c} dt_1 \cdots \int_0^{t_{n-1}} dt_n \left\| \prod_{j=1}^n \tilde{H}_0(t_j) \right\| \\ &\stackrel{(2)}{\leq} \int_0^{\tau_c} dt_1 \cdots \int_0^{t_{n-1}} dt_n \prod_{j=1}^n \|\tilde{H}_0(t_j)\| \\ &\stackrel{(3)}{=} \int_0^{\tau_c} dt_1 \cdots \int_0^{t_{n-1}} dt_n \prod_{j=1}^n \|H_0\| \\ &= \frac{\tau_c^n}{n!} \|H_0\|^n \stackrel{(1)}{\leq} \frac{\tau_c^n}{n!} (J + \beta)^n. \end{aligned} \quad (\text{A11})$$

As a result, Eq. (A10) achieves the upper-bound

$$D \leq \frac{1}{\sqrt{2}} \sum_{n=1}^{\infty} \frac{\tau_c^n}{n!} (J + \beta)^n = \frac{1}{\sqrt{2}} \left(e^{\tau_c(J+\beta)} - 1 \right). \quad (\text{A12})$$

Let us now consider the case of N th order error suppression, where all $\tilde{U}_0^{(n)}(\tau_c)$ for $n \leq N$ vanish. The bound then truncates to

$$D \leq \frac{1}{\sqrt{2}} \sum_{n=N+1}^{\infty} \frac{\tau_c^n}{n!} (J + \beta)^n, \quad (\text{A13})$$

which scalings accordingly as

$$D \lesssim \mathcal{O}[\tau_c^{N+1} (J + \beta)^{N+1}], \quad (\text{A14})$$

when $J\tau_c \ll 1$ and $\beta\tau_c \ll 1$ are satisfied.

Appendix B: Extracting effective error Hamiltonian scaling numerically

In Sec. IV A, we discussed the scaling of the distance measure D for each optimal sequence in the ideal pulse limit without direct calculation of the effective Hamiltonian. We obtain this scaling by assuming that D has the form

$$D \sim \mathcal{O}(J^{n_J} \beta^{n_\beta} \tau_d^{N+1}), \quad (\text{B1})$$

where N is the decoupling order of the sequence and $n_J + n_\beta = N + 1$. First, the decoupling order is determined by examining $\log_{10} D$ as a function of τ_d , as this quantity scales linearly in τ_d with a slope of $N + 1$. The scaling of τ_d is only dependent upon the decoupling order N , and therefore is independent of the relative magnitudes of J and β . The only constraint we consider is $J\tau_d \ll 1$ and $\beta\tau_d \ll 1$ in order to satisfy the condition $\|H'_{\text{eff}}\tau_d\| \ll 1$ discussed in Sec. II A for effective error suppression.

The scaling of the remaining quantities, J and β , is determined for each relevant parameter regime ($J \ll \beta$ and $J \gg \beta$) by analyzing $\log_{10} D$ as a function of $J\tau_d$ by varying J for fixed τ_d and β . The logarithm of the performance measure can again be expected to scale linearly in $J\tau_d$, now with a slope of n_J . The value of n_J will ultimately depend on the magnitude of J relative to β , however in either case n_J is well-defined. The scaling of β is now determined from $n_J + n_\beta = N + 1$, where n_β is the only unknown quantity. Note that this method can be easily extended to include finite-width pulses, flip-angle errors, or both, by assuming

$$D \sim \mathcal{O}(J^{n_J} \beta^{n_\beta} \tau_p^{n_p} \tau_d^{n_d}), \quad (\text{B2a})$$

$$D \sim \mathcal{O}(\epsilon^{n_\epsilon} J^{n_J} \beta^{n_\beta} \tau_d^{N+1}), \quad (\text{B2b})$$

$$D \sim \mathcal{O}(\epsilon^{n_\epsilon} J^{n_J} \beta^{n_\beta} \tau_p^{n_p} \tau_d^{n_d}), \quad (\text{B2c})$$

respectively, where $n_p + n_d = n_J + n_\beta$.